Assessing Telecommunication Service Availability Risks for Crisis Organisations

Assessing Telecommunication Service Availability Risks for Crisis Organisations

Eelco Vriezekolk

Enschede, The Netherlands, 2016

Ph.D. dissertation committee:

Chairman and secretary: prof.dr. P.M.G. Apers
Universiteit Twente, EWI

Promotors: prof.dr. R.J. Wieringa
Universiteit Twente, EWI
prof.dr. S. Etalle
Universiteit Twente, EWI

Members: prof.dr. J. van Hillegersberg
Universiteit Twente, BMS
prof.dr.ir. A. Pras
Universiteit Twente, EWI
prof.dr. K. Schneider
University of Hannover
prof.dr.ir. J.C. Wortmann
Rijksuniversiteit Groningen
prof.dr. R. Breu
University of Innsbruck

CTIT Ph.D. Thesis Series No. 16-393
Centre for Telematics and Information Technology
P.O. Box 217, 7500 AE Enschede, The Netherlands

SIKS Dissertation Series No. 2016-32
This research has been carried out under the auspices of SIKS, the
Dutch research School for Information and Knowledge Systems.

This research was generously supported by
Radiocommunications Agency Netherlands.

ISSN 1381-3617
ISBN 978-90-365-4141-1
DOI 10.3990/1.9789036541411
<http://dx.doi.org/10.3990/1.9789036541411>

Copyright © 2016, Eelco Vriezeekolk, The Netherlands

ASSESSING TELECOMMUNICATION SERVICE AVAILABILITY RISKS FOR CRISIS ORGANISATIONS

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
prof.dr. H. Brinksma,
on account of the decision of the graduation committee,
to be publicly defended
on Thursday 14th of July 2016 at 12.45h

by
Eelco Vriezekolk

born on 12 August 1966
in Hengelo (O), The Netherlands

This thesis has been approved by:

prof.dr. R.J. Wieringa promotor

prof.dr. S. Etalle promotor

Samenvatting

De zorg voor de veiligheid van haar burgers is de eerste en meest belangrijke taak van de overheid. Crisisorganisaties zijn publieke organisaties die de samenleving helpen tijdens veiligheidsincidenten, en die incidenten helpen voorkomen en beperken. Crisisorganisaties bestaan uit hulpverleners (politie, brandweer, medische hulp, etc), crisis-coördinatiecentra en bestuurders. Een vereiste voor efficiënte crisisbeheersing is dat informatie snel gedeeld wordt, om de hulpverlening te coördineren en slagvaardig besluiten te kunnen nemen. Moderne crisisorganisaties zijn daarom afhankelijk van telecommunicatiediensten, met name wanneer zij net-centrisch werken. Als tijdens een incident telecommunicatiediensten niet beschikbaar zijn, zal de schade toenemen en kunnen mensen sterven. Om goed voorbereid te zijn, moeten crisisorganisaties hun risico's van uitval van telecomdiensten weten: zij moeten een risicobeoordeling uitvoeren.

Risicobeoordeling in dit domein is een uitdaging. Telecomdiensten zijn samengesteld uit netwerken en diensten van meerdere, onafhankelijke en concurrerende bedrijven. Dat maakt het erg moeilijk om betrouwbare informatie over het netwerk te verkrijgen. Zelfs als alle informatie beschikbaar zou zijn, dan nog is een risicomodel dat alle fysieke componenten bevat moeilijk te construeren omdat het bovenmatig complex zou zijn. Telecommunicatienetwerken veranderen voortdurend terwijl er relatief weinig serieuze incidenten gebeuren, waardoor gegevens voor een betekenisvolle statistische analyse van incidenten moeilijk te verkrijgen zijn. Ten slotte kunnen in dit domein de risicobeoordelingen niet alleen worden gebaseerd op technologische factoren; de belangen en voorkeuren van de samenleving zijn ook relevant. Om de risico's van uitval van telecomdiensten te beoordelen is een risicobeoordelingsmethodologie nodig die effectief en efficiënt met deze moeilijkheden kan omgaan. Uit het onderzoek is geen reeds bestaande risicobeoordelingsmethode naar voren gekomen die aan deze eisen voldoet.

Om te zorgen dat crisisorganisaties wel een risicobeoordeling kunnen uitvoeren, hebben wij een methode ontworpen, genaamd RASTER, die is afgestemd op dit domein en de uitdagingen ervan. RASTER is op drie principes gebaseerd:

- Risicobeoordeling vereist samenwerking tussen experts met diverse professionele achtergronden. RASTER moet voor allen makkelijk te gebruiken zijn, en samenwerking faciliteren.
- Betrouwbare statistische informatie over fouten en uitval is vaak niet beschikbaar, waardoor kwalitatieve expert-oordelen nodig zijn.

- Risico-prioriteiten moeten zijn gefundeerd op objectieve feiten, maar moeten ook rekening houden met voorkeuren van betrokkenen.

RASTER tekent en gebruikt diagrammen van telecommunicatiediensten. Deze diagrammen werken als een gemeenschappelijke grafische taal tussen experts. Diagrammen hoeven niet vooraf in detail te worden opgesteld; details worden toegevoegd als en wanneer deze nodig zijn. Dit verklaart de naam van RASTER: Risk Assessment by Stepwise Refinement (Risicobeoordeling door Stapsgewijze Verfijning).

Dit proefschrift beschrijft een ontwerpwetenschapsaanpak voor de ontwikkeling van de RASTER methode, vanaf de eerste specificatie en ontwerp via diverse verbeterstappen tot aan zijn uiteindelijke vorm. Ontwerpwetenschap creëert artefacten die kunnen worden gebruikt om praktische problemen te behandelen binnen een bepaalde context. In het maken van RASTER hebben we telkens afgewisseld tussen het beantwoorden van kennisvragen en het oplossen van praktische problemen. In deze ontwikkeling zijn meerdere lab-experimenten gehouden om de bruikbaarheid en betrouwbaarheid van de methode te valideren. Er zijn twee veldtesten gehouden, waarin de auteur de nieuwe methode heeft toegepast om praktische vraagstukken bij twee crisisorganisaties te helpen oplossen. Dit proefschrift beschrijft daarom theoretisch onderzoek, experimenten in laboratorium- en veld-omgevingen, en technisch actieonderzoek. Met uitzondering van het laatste experiment hebben alle experimenten geleid tot verbeteringen aan de methode.

Dit onderzoek is in Nederland uitgevoerd, maar de resultaten zijn niet afhankelijk van Nederlandse crisisstructuren en zouden ook toepasbaar moeten zijn in andere landen.

Met behulp van RASTER kunnen crisisorganisatie nu de uitvalsrisico's van de door hen gebruikte telecomdiensten ontdekken, analyseren en prioriteren. Crisisorganisaties kunnen beter voorbereid zijn, en zo bijdragen aan een veiliger samenleving.

De auteur bedankt alle deelnemers aan dit onderzoek, in het bijzonder de experts van Agentschap Telecom, de Veiligheidsregio Groningen en Waterschap Hunze en Aa's, en betuigt zijn oprechte dankbaarheid aan Agentschap Telecom voor het mogelijk maken van dit onderzoek.

Abstract

Protecting the safety of its citizens is the first and foremost responsibility of government. Crisis organisations are public organisations that assist society when safety incidents occur, and help to prevent and limit incidents. Crisis organisations include first responders (police, fire services, emergency medical care, etc), crisis coordination centres and decision makers. Their effective incident management requires rapid information sharing to coordinate operations and expedite decision making. Modern crisis organisations therefore depend on telecommunication services, especially since many have adopted net-centric operations. When telecommunication services are unavailable during an incident, damage will increase and people may die. In order not to be caught unprepared, crisis organisations must know their telecom service availability risks: they need to perform a risk assessment.

Risk assessment is challenging in this domain. Telecom services are composed of networks and services of many independent, competing companies which makes it very hard to obtain reliable information about the network. Even if complete information were available, a risk model showing all physical components is difficult to construct because it would be excessively complex. Telecom networks change continuously with serious incidents being relatively rare, which means that data for meaningful statistical analysis of incidents is hard to obtain. Lastly, in this domain risk assessment cannot be based on technological factors only; the priorities and preferences of society are relevant as well. To assess telecom service availability risks for crisis organisations, a risk assessment methodology is needed that handles these complications efficiently and effectively. This research did not identify an existing risk assessment method matching these requirements.

To make it possible for crisis organisations to perform a risk assessment, we have developed a method called RASTER that is tailored for this domain and its challenges. RASTER is based on three principles:

- Risk assessment requires collaboration among experts from diverse professional backgrounds. RASTER should be easy to use by all, and facilitate teamwork.
- Reliable statistical information on faults and failures is often unavailable, making qualitative expert judgment necessary.
- Risk prioritisation should be based on objective facts, but should take stakeholder preferences into account as well.

RASTER creates and uses diagrams of telecommunication services. These dia-

grams function as a common graphical language among experts. Diagrams need not be specified in detail in advance; details can be added as and when necessary. This explains the RASTER name: Risk Assessment by Stepwise Refinement.

This thesis describes a design science approach to the development of the RASTER method, from its first specification and design through several improvement steps to its final form. Design science creates artifacts that can be used to treat practical problems within some context. In creating RASTER, we continuously iterate between answering knowledge questions and solving practical problems. As part of this development several lab experiments were held to validate the usability and reliability of the method. Two field tests were held in which the author applied the new method to help solve practical problems at two crisis organisations. This thesis therefore describes theoretical research, experiments in lab and field settings, as well as technical action research. After each experiment except the last one, improvements were made to the design.

The research was carried out in the Netherlands, but the results does not depend on Dutch crisis structures and should be applicable to other countries as well.

Using RASTER, crisis organisations can now discover, analyse and prioritise the availability risks of the telecommunication services that they use. Crisis organisations can be better prepared, helping society to be safer.

The author thanks all participants in this research, in particular the experts at Agentschap Telecom, Safety Region Groningen and Waterschap Hunze en Aa's, and sincerely expresses his gratitude to Agentschap Telecom for making this research possible.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Crisis organisations	2
1.2.1	Incident, crisis, disaster	3
1.2.2	The Dutch government's approach to crisis management	4
1.2.3	Safety chain	5
1.2.4	Network centric operations	6
1.3	Problem statement	7
1.4	Research methodology and contribution	8
1.5	Thesis outline and publications	10
2	Risk and Risk Assessment in Telecommunication	13
2.1	Terminology	13
2.1.1	Sources of terminology	14
2.1.2	Risk target, environment	16
2.1.3	Asset, stakeholder	16
2.1.4	Hazard, safety	17
2.1.5	Failure, Vulnerability	19
2.2	Uncertainty	20
2.3	Risk	22
2.3.1	Facts and values	23
2.3.2	Risk perception and communication	24
2.4	Risk management	24
2.4.1	Risk identification	26
2.4.2	Risk analysis	26
2.4.3	Risk evaluation	28
2.4.4	Risk treatment	29
2.5	Conclusion	29
3	Current Practice and Theory	31
3.1	Overview	31
3.1.1	Engineering	31
3.1.2	Crisis management	34
3.1.3	Information systems	35

3.2	Selected methods and standards	36
3.2.1	Engineering	36
3.2.2	Crisis management	40
3.2.3	Information technology	44
3.3	Discussion	46
4	Requirements	49
4.1	A risk management framework	50
4.1.1	Arguments for risk estimates	50
4.1.2	Trade-offs in decision making	52
4.1.3	List of risk factors	53
4.1.4	Combining different meanings of risk	56
4.1.5	Defining adequacy	57
4.2	Cases	57
4.2.1	Case Health risks of electromagnetic fields	57
4.2.2	Case Triple play	59
4.2.3	Case C2000	59
4.2.4	Discussion	60
4.3	Requirements of risk assessment methodology	61
4.3.1	Challenges from crisis management	61
4.3.2	Challenges from telecommunications	61
4.3.3	Initial requirements	63
4.4	Current methods versus the requirements	65
4.5	Conclusion	66
5	Initial Design of the Raster Method	67
5.1	Telecom service models	67
5.2	Risk analysis in telecom models	71
5.2.1	Use of risk factors	71
5.2.2	Evaluation of vulnerability score	73
5.2.3	Evaluation of overall vulnerability level	75
5.2.4	Telecom service risk evaluation	76
5.3	Execution of the Raster method	76
5.4	Discussion	78
5.4.1	Limitations of telecom service models	78
5.4.2	Limitations of the Raster method	79
5.5	Design validation	80
6	Making Raster Work	83
6.1	Research method	83

6.1.1	Research questions	83
6.1.2	Case description	84
6.2	Case study execution	84
6.2.1	Questionnaire and final interview	87
6.3	Results and discussion	88
6.3.1	Case study execution	89
6.3.2	Questionnaire and final interview results	90
6.3.3	Research questions	92
6.4	Lessons learned	93
6.5	Design improvements	94
6.5.1	Conceptual model	94
6.5.2	Tool support	95
6.5.3	Telecom service diagrams	95
6.5.4	Risk analysis	96
6.5.5	Execution of the method	99
6.6	Conclusion	100
7	Testing Raster's Reliability	101
7.1	Introduction	101
7.2	Background and related work	102
7.3	Our approach to testing reliability of a method	103
7.3.1	Controlling variation	103
7.3.2	Analysis of measurements on the results of the method	105
7.4	Research method	107
7.4.1	Experiment design	107
7.4.2	Using our approach to testing reliability	109
7.5	Results	110
7.5.1	Scoring results	110
7.5.2	Exit questionnaire	110
7.5.3	Implications	114
7.6	Design improvements	114
7.7	Conclusion	118
8	First Field Test	121
8.1	Setting and method	121
8.1.1	Research questions	122
8.1.2	Method	123
8.2	Results	124
8.3	Discussion	130

8.3.1	Research questions	131
8.4	Design improvements	132
8.5	Conclusion	133
9	Second Field Test	135
9.1	Setting and method	135
9.2	Results	137
9.3	Discussion and comparison	143
9.3.1	The Raster method	143
9.3.2	Research questions	147
9.4	Peer review	148
9.5	Lessons learned	152
9.6	Conclusion	153
10	Conclusions	155
10.1	Summary	155
10.2	Review of research questions	156
10.3	Generalisation and limitations	159
10.4	Outlook and further research	162
10.5	Final remarks	162
Appendices		
A	Krippendorff's alpha	165
A.1	Computing alpha	165
A.2	Alpha over subsets	167
A.3	Example	167
B	Questionnaires	171
B.1	Improving the initial design	171
B.2	Reliability experiment	183
B.3	Field test questionnaires	197
C	Raster software tool	207
D	Raster Method – application manual	209
Author publications		251
References		253
Index		263

Introduction

1.1 Motivation

In November 2009, a failure in a switch in Amsterdam ultimately results in suspension of a rail service in the city of Utrecht [118]. Although there is no direct relation between the failure and the rail service, the affected GSM network is used for communication between train drivers and the traffic control centre in case of emergencies. Since communication is no longer possible, the transport company decides to suspend transport as a precautionary measure.

During morning rush hour in May 2011, traffic jams on the A2 highway in the Netherlands are much higher than normal. The electronic traffic signs over the highway lanes display incorrect or conflicting information and cannot be controlled remotely [101]. The problems are caused by an underground cable that is damaged during construction work.

On 27 July 2011, a telecommunications exchange in Rotterdam fails [74]. As a result, metro and other public transport services are suspended, automatic fire alarms go offline, and various emergency communication systems behave erratically. High-risk cargo unloading operations in the port of Rotterdam have to be suspended, and baggage handling at a nearby local airport has to be performed manually as the automatic system stops functioning.

In the night of 20–21 June 2012 routine scheduled maintenance is performed on a fiber optic cable. This cable is used within the 1-1-2 system, the national emergency telephone number. Unknown to all involved parties is that the redundant backup line happens to be out of order. As a result, with both the main cable and the backup line unavailable, half of all emergency calls go unanswered. The incident lasts over 6 hours, and during this period two people die. A subsequent investigation fails to show a causal relationship between the outage and their deaths.

These four cases are just a few examples of society's dependence on telecommunications. Telecommunications, both wired or wireless, are essential to the undisturbed functioning of society. Disruptions can have unexpected consequences.

Figure 1.1 – Crisis organisations in action.

© Pim Velthuisen

One of the problems is that we cannot, in general, predict what activities will be affected when a particular telecommunications service fails. Because telecommunication outages are relatively rare there is little awareness of the fact that telecom can fail, and many stakeholders have a fairly naive understanding of the possible causes and mechanisms of failures. A proper understanding of failure modes is difficult to obtain, because telecommunication services are increasingly complex. Radiocommunications Agency Netherlands, a Dutch inspectorate and policy implementation unit for the telecommunications sector, indicates that the combination of high dependency and incomplete understanding of availability risks is a serious threat to society [133].

The challenge for telecom users is to discover all telecommunication services on which they depend, and to assess the availability risks to these services.

1.2 Crisis organisations

This challenge is an issue for any organisation, but more so for organisations that manage critical infrastructures or critical public functions. Crisis organisations are an example of these. Crisis organisations respond to large accidents, such as fires, explosions, flooding, road and air traffic accidents, outbreaks of infectious diseases, bomb threats, and other natural or human-made disasters (see Figure 1.1). When operations of crisis organisations are disrupted through some failure of telecommunications, consequences will be serious. Lives of citizens and emergency responders may be at risk.

The research problem for this thesis will be formally stated in the next section, but informally the issue is how crisis organisations can reduce the risk of telecom services becoming unavailable. Since crisis organisations are central to this thesis, the remainder of this section provides background information on crises, crisis management, and how crisis management is organised in the Netherlands.

1.2.1 Incident, crisis, disaster

The terms ‘incident’, ‘crisis’, and ‘disaster’ all indicate ‘a hazard that becomes reality’. They differ in the type and magnitude of the resulting consequences, and in the coping capabilities of the organisation and its stakeholders. As with many risk and crisis related terms, there is little agreement on the exact meaning of terms. This section stays close to the meaning of the terms as in Dutch regulations and practice of public crisis organisations [112, 116]. ‘Hazard’ will be defined in Section 2.1.4.

Any hazardous event that can be handled by routine procedures is termed an *incident*. Conventional traffic accidents are examples of incidents. If the incident requires an urgent response, it is termed an *emergency*. For example, the Dutch emergency number 1-1-2 is promoted with the slogan “when every second counts”. If the incident is potentially harmful to critical functions of society it is termed a *crisis*. The collapse of the banking system in 2009 is an example of a (financial) crisis. If a crisis, especially when its effect is mostly physical, cannot be handled by the local community without outside help, it is termed a *disaster*. The term *catastrophe* is used to indicate that the disaster caused extreme damage. Crises and disasters require a response that is coordinated among different organisations. Most disasters will at some stage require an urgent response, and ‘disaster’ and ‘emergency’ therefore overlap. Crises in general do not have to be emergencies. For example, climate change can be termed an (ecological) crisis that needs to be addressed in the coming years. It is not, however, urgent in the sense that it requires sirens and speeding emergency vehicles.

Decision making for disasters is difficult because essential information is often unavailable and decisions have to be made under severe time pressure. It is therefore highly stressful for the decision makers involved [92], [112, Chapter 20]. Decision making for crises is often hampered by uncertainties about the causes and effects, and about the effectiveness of actions. Also, there are often conflicting goals, conflicting opinions, and limited resources. Global climate change is a case in point.

What starts as an incident can evolve into a crisis. But for many past crises it is difficult to pinpoint a clear originating incident. Rather there may be a diffuse confluence of events that slowly regresses into a full-blown crisis. On the other hand, for some types of events it is immediately clear that routine procedures

Figure 1.2 – Relations over time between hazard, incident, crisis, and restoration, and their management activities.

will be insufficient, such as the 2011 tsunami disaster in Japan. Figure 1.2 shows the relationship between hazards, incidents, and crises (including disasters and catastrophes).

Crisis management involves the actions necessary to respond to incidents, and the preparation for such actions. Crisis management encloses incident management as a special case. The term ‘crisis management’ is therefore often used to cover both activities, not just crisis management proper. In this thesis the term is commonly used in this wider meaning.

The goal of crisis management is then to prevent incidents from happening, reduce their impact, and to recover from incidents quickly and with the minimum amount of damage [79].

1.2.2 The Dutch government’s approach to crisis management

Before 2010 the Netherlands had three acts governing public crisis management. In 2010 these were replaced and renewed by a single act: the Safety Regions Act. The Explanatory Memorandum of this act states [116]:

“Taking care of safety is a fundamental duty of government. Citizens have the right to a government that takes whatever measures can reasonably be demanded to create a safe environment.”

A Safety Region is an organisation responsible for fire services, emergency medical care and crisis management and response. Safety Regions serve areas with an average population of 650 thousand; there are 25 Safety Regions in the Netherlands.

The Safety Regions Act uses the term ‘disaster’ (in Dutch: *ramp*), which it defines as “a severe accident or other event where the lives and health of many people, the environment or significant material interests are seriously harmed or threatened and where the coordinated deployment of services and organisations of several disciplines is required to remove the threat or reduce the damaging impact”. A disaster denotes a “classical crisis” [116], in which the event and its consequences are clear and limited in period and location. In recent years government noticed a shift towards more complex crises, where there is no clear cause or a combination of causes, consequences are diffuse and spread out in geography and time, and where conflicting opinions exist on the

<i>Stage</i>	<i>Definition</i>
Pro-action	Eliminating structural causes of accidents and disasters to prevent them from happening in the first place (e.g. by proscribing building in flood-prone areas)
Prevention	Taking measures beforehand that aim to prevent accidents and disasters, and limit the consequences in case such events do occur (e.g. by building dikes and storm surge barriers)
Preparation	Taking measures to ensure sufficient preparation to deal with accidents and disasters in case they happen (e.g. contingency planning)
Response	Actually dealing with accidents and disasters (e.g. response teams)
Recovery	All activities that lead to rapid recovery from the consequences of accidents and disasters, and ensuring that all those affected can return to 'normal' and recover their equilibrium

Figure 1.3 – The stages in the safety chain (from [113]).

best course of action. The Safety Regions Act uses the term 'crisis' for these complex situations, which it defines as "a situation in which a critical interest of society is harmed or is under threat".

1.2.3 Safety chain

The Dutch government presents and organises activities and decisions for crisis management in a scheme that it calls the 'safety chain'. This model was first introduced in the first Safety Report in 1993 [113]. The safety chain is an extension of a model used by the United States Federal Emergency Management Agency (FEMA) [152]. The difference between the two models is that the safety chain separates pro-action from prevention. This split causes some ambiguity on the placement of activities aimed at reducing the likelihood of hazards and threats. The definitions and explanation in the Safety Report are unclear on this point. This split was motivated by the desire to "push" the fire services into a broader interpretation of prevention. Policy makers, at the time, were of the opinion that fire services were focusing too much on fire prevention. The designation of pro-action was believed to help fire services take a broader view on prevention [62, p.30]. Pro-action is limited to those activities that remove risks entirely; risk reduction activities are considered to belong to the prevention stage of the safety chain. *Figure 1.3* shows the stages in the safety chain.

Pro-active measures include zoning laws and restrictions on routes for dangerous transports. For example, transport of certain hazardous substances is not allowed in tunnels. Pro-active measures are totally effective, as the risk is removed entirely. However, pro-action often depends on physical separation or termination of hazardous activities. Within the limited area that the Netherlands offer, this course of action is often economically costly [112]. This leaves preventive measures as the next recourse. Preventive measures include various

technical means to e.g. reduce fire hazards. The costs of preventive measures are typically borne by the private sector, whereas the costs of pro-active measures are typically borne by government or by society as a whole [112]. One notable counter example here is the construction of dikes: a preventive measure (dikes do not remove the risk of flooding entirely) built at government's expense.

Measures that belong to the preparation stage include planning, training, exercises, and warning systems. These measures do little to reduce the likelihood or potential impact of hazards, but do help in reducing the consequences.

The response stage is the heart of crisis management. Fire fighting, search and rescue parties and forced evacuations all belong to this stage. Improvisation is often essential in this stage, and legislation makes allowance for emergency authority of public officials.

Recovery is the last, but also often the first stage in the safety chain. In addition to restoration and reconstruction, the recovery stage is also used to rethink risk assessments, policies, and procedures.

The stages in the safety chain are more than a simple collection of measures, activities and decisions. Each stage should be aligned with the preceding and following stage. The value of the safety chain lies therein that it enables a discourse about the gaps between the stages, and that it provides arguments for trading off resource allocation between the stages [112].

Crisis management requires a multidisciplinary approach. The Dutch crisis management organisations recognise five major partners and disciplines, often indicated by their typical colour: fire and emergency services (red), medical care (white), police (blue), national defence (green), and local government (orange). Each discipline adds its own knowledge and capabilities to the crisis management pool. The Safety Regions Act addresses all of these five disciplines.

1.2.4 Network centric operations

Since many actors from several disciplines are involved in crisis management, information sharing is critical to operations. Nowadays, information sharing requires the use of telecommunications and information technology. The military first recognised the benefits of centralised information collection, analysis, and dissemination. Their concept of 'network centric warfare' has been around since the end of last century [3]. It is itself an effect of the profound changes brought about by the rise of information technology, exemplified by the popular rise of the Internet during the 1990's. The central idea behind network centric warfare is that existing command and control structures as well as the operational units are aided by improved and shared situational awareness. Geographic information systems and other IT technologies enable the rapid visualisation, retrieval, and dissemination of information. This improves the quality and timeliness of decision making during crisis situations. Network centric warfare requires the rapid collection and processing of information from many different sources

(observations by people as well as various sensors), for which Internet-age information technology is deemed essential.

The ideas behind network centric warfare were quickly adopted by crisis management professionals. Improved and shared situational awareness are of benefit to crisis management as well [161]. In the Netherlands, early trials with network centric operations were started in 2005. These trials were successful, and the regulations accompanying the Safety Regions Act now mandate the use of network centric operations in all safety regions [115, article 2.4]. Of course, network centric crisis management is heavily dependent on telecommunications.

1.3 Problem statement

This research explores the dependencies of crisis organisations on availability of telecommunication services and the vulnerabilities of these services. Its aim is to present a method for analysing the dependencies, and for assessing the risks that crisis organisations face in case of telecommunication failures. This research has an overarching practical aim: not so much to further theory or the state of the art, but to solve a practical problem that actually exists in current practice. Therefore, where a choice is unavoidable, usability by the target audience supersedes analytical meticulousness. Where theoretical research would have stopped short at validation in industry practice, this research considers multiple field tests of the method as the culminating goal. The research therefore would not have been possible without extensive practical experience, nor without a suitable network within the target community of crisis organisations.

We define the following as our research problem:

How can a crisis organisation discover the availability risks to the telecommunication services on which it depends?

This research does not directly assess the mitigating actions that crisis organisations can take to reduce those risks. In some cases the most effective countermeasures are straight forward. For example, if short-term power failures are a high risk, then installation of a battery-powered backup supply will be effective. In other cases countermeasures involve redesigning telecommunication infrastructures.

The scope of this research problem is limited to public crisis organisations and the telecommunications services they use. Generalisation to other organisations may be possible, and in some cases even obvious, but is not a direct research goal. The final chapter addresses to what extent this research can be generalised to other types of organisations, and to infrastructures other than telecommunications.

To solve the research problem we answer three main research questions. In RQ1 we address the current state of the art in risk assessment and current

practice in our problem context. We further investigate the research problem, and derive requirements for risk assessment of telecom service availability. RQ1 is a knowledge problem, which means that we collect knowledge about the problem without solving it. We investigate the problem within its context. To do so, RQ1 is decomposed into the following sub-questions.

RQ1: Are current risk assessment methods able to adequately assess availability risks in our problem context?

RQ1-a: What existing risk assessment methods can be applied to telecommunication services and their use in crisis organisations, and what are their properties?

RQ1-b: What risk assessment methods are currently used by crisis organisations?

RQ1-c: What are the requirements for risk assessment methods in our problem context?

RQ1-d: What risk assessment methods do match those requirements?

In RQ2 we address the design of a new risk assessment method to fit the requirements from RQ1-c. This is a design problem, which means that the solution calls for the creation of an artifact. The artifact here is a risk assessment method.

RQ2: How can we design a risk assessment method for availability risks in our problem context?

In RQ3 we take this artifact and validate it against reliability, correctness, and the specific requirements derived from question RQ1-c. RQ3 is again a knowledge problem; we collect information about the method. The validation consists of the following subquestions.

RQ3: What is the contribution of our new risk assessment method?

RQ3-a Is the new risk assessment method feasible: can the method be performed in practice?

RQ3-b Is the new risk assessment method reliable: can the method be repeated with comparable results?

RQ3-c Is the new risk assessment method an improvement over current methods?

1.4 Research methodology and contribution

To develop this method we use the design science methodology proposed by Wieringa [165]. Design science creates an artifact by designing and investigating it in a context. The goal of the artifact is to make changes to the world: to

Figure 1.4 – The engineering cycle, used to solve design problems (from [165]).

improve upon some problem. Each artifact is therefore a treatment specification. The term ‘treatment’ instead of ‘solution’ expresses that the changes may not be fully effective, and may introduce new problems. The artifact in this thesis is a method; in particular it is a method to assess availability risks, and the context here is that of crisis organisations.

We create our method using the engineering cycle, as in *Figure 1.4*. The engineering cycle contains two kinds of activities: answering knowledge questions, and solving practical problems. In the diagram these are distinguished by a question mark and exclamation mark respectively. The engineering cycle is not necessarily performed in a single cycle. Treatment validation may yield answers that lead to treatment redesign, and the steps of design and validation may have to be performed several times. Because of this, questions RQ2 and RQ3 are not answered in a linear fashion. Instead, we perform a series of tests whereby we create a design, validate it in the lab or in the field to learn about its properties, and improve the design based on our findings. We have performed this improvement cycle a number of times.

Our research aims to create a method that can be used by professionals in practical situations when time and effort required to execute the method are limited. This means that we focus on obtaining actionable recommendations for reducing availability risks, and on practical usability in uncontrolled circumstances. A large part of the research was spent on field tests. In a field test, external influences are unpredictable and uncontrollable, unlike in a lab environment where all procedural aspects can be carefully controlled.

The first validation experiment tested whether the method could be performed from beginning to end, with acceptable effort and within acceptable time. This experiment was carried out at Radiocommunications Agency Netherlands, using internal experts and the internal crisis organisation as a subject. The need for tool support was one of its outcomes. As a consequence, work was started on a prototype of a tool to create diagrams of telecommunication services and to record and analyse availability risks. Several hundred hours have been spent

on trying out various options and on adding necessary features. By the end of this research, the prototype had evolved into a medium-sized, full-featured software program.

To validate the reliability of the method, we conducted an experiment in which six groups performed the core analysis part of the method independently and in parallel. For this experiment we acquired student volunteers, developed an artificial test case and training material, and held experiment sessions at two universities. Planning a volunteer experiment and recruiting participants takes a lot of time; it took several months from initiation to the start of the first experiment session. This experiment, too, resulted in improvements to the design.

Finally, the method was tested in uncontrolled environments in field tests at Dutch crisis organisations. In these tests, the author applied the method to assist an organisation in solving a practical problem in a professional, uncontrolled environment. Three Safety Regions and a Water Board were approached. Eventually two host organisations were found that were both amenable to the need for risk assessment and able to participate in the test. The first field test was expedited through previous contacts between the host organisation and Radiocommunications Agency Netherlands in its role as a government agency. The second field test resulted from the recommendation by an enthusiastic participant in the first test. Even so, it took five months for the host organisation to be ready for the second field test. Both field tests required informal and formal meetings with the management of the crisis organisation to obtain approval. Subsequently, five to ten work sessions with experts from the crisis organisation were needed to complete all the risk assessments. The projects' basic results, if printed, would be 70 to 80 pages each. The two field tests resulted in a 30- to 40-page internal report summarising the assessments, describing the top-priority availability risks and recommendations for risk treatment. These reports are confidential, because they explicitly describe weaknesses in internal telecommunication systems. The results were reported to management, and most treatment recommendations were implemented in practice.

This research therefore combines theoretical research, experiments in lab and field settings, as well as technical action research, using design science as its theoretical framework.

1.5 Thesis outline and publications

The chapters in this thesis are based on previously published conference papers; the full list of the Author Publications can be found in the References section (page 251). These papers have been edited to make them suitable for reprinting in this thesis: introductions have been condensed or removed to avoid unnecessary repetition, and sections of papers on related research have been moved into

Figure 1.5 – Outline of this thesis, showing the chapters and the research questions addressed in them, and their place in the engineering cycle.

their own chapter, where necessary. In a few places original material that had to be dropped in order to satisfy proceedings page count constraints has been restored. References to these author publications is given in the introduction to each chapter. This structure is depicted in *Figure 1.5*.

Chapter 2 contains definitions and background information on risk, risk assessment, crises, and crisis management. The contents of this chapter have not been published separately.

Chapter 3 reviews current practice; in this chapter we suggest that existing methods are not sufficient for our needs. Addresses research questions RQ1-a and RQ1-b. Part of this chapter are taken from Author Publication 3.

Chapter 4 develops and formulates the requirements for risk assessment methods in our domain, and justifies the design of a new risk assessment method. Addresses research questions RQ1-c and RQ1-d. Requirements have been first formulated in Author Publication 5; Author Publication 4 examines risk factors.

Chapter 5 describes the initial design of our new risk assessment method, called RASTER. Addresses research question RQ2. This chapter is based on Author Publication 5 and an early version of the RASTER application manual.

Chapter 6 validates the feasibility of RASTER. In a number of improvement steps the method is evolved until it is possible to perform the method within the limits of acceptable time and effort. Addresses research question RQ3-a. This chapter is based on Author Publication 3; the section describing design

improvements has not been published previously.

Chapter 7 validates the reliability of the method. Addresses research question RQ3-b. This chapter is based on Author Publication 2, and parts of an unreviewed technical report (Author Publication 10).

Chapters 8 & 9 describe the field tests. Address research questions RQ3-a and RQ3-b. The first field test has been published as Author Publication 1; the second as an unreviewed technical report (Author Publication 8). Only the first validation led to improvements to the method.

Chapter 10 answers our research questions and describes implications for practice and for further research. Addresses research question RQ3-c, and answers RQ1, RQ2, and RQ3. This chapter is not based on previous publications.

Risk and Risk Assessment in Telecommunication

This chapter explores the broader subject of risk and risk assessment, and the terms used. It has two goals. First, it discusses key terms for which alternative terms or multiple definitions are in use (Section 2.1). This section also explains why different ones are sometimes needed, and chooses which ones will be used in the remainder of this thesis (marked by notes in the margin of the text). Secondly, the chapter describes and discusses three central concepts: uncertainty, risk and risk management (Sections 2.2 to 2.4).

2.1 Terminology

A multitude of definitions is currently in use for key terms such as risk, hazard, vulnerability, and others. There is no single universally accepted definition of risk and its aspects, and the potential for misunderstandings is high. For example, Hansson [58] mentions four different meanings of the term ‘risk’:

1. *an unwanted event*, as in “lung cancer is one of the major risks that affect smokers”;
2. *the cause of that event*, as in “smoking is a health risk”;
3. *the probability of the event*, as in “the risk that a bridge will collapse”; and
4. *the statistical expectation value of the event*, as used in gambling and financial services.

Aven and Renn [7] mention ten common definitions, and propose an eleventh one. In another publication, Thywissen [156] collected no fewer than 36 different descriptions of the term ‘vulnerability’.

Not only do different authors use different definitions, either explicitly or implicitly, but each academic discipline also has its own preferences and subtle differences. To understand the meaning of a text, it is therefore important to understand the academic background of the authors, as well as their personal nuances. For example, Christensen et al. [21] observed that the health sciences

typically emphasise the uncertainty (the probability) of threats (number 3 in the list above), and less so their expected impact (number 1 in the list), as in other disciplines. In security research, the term ‘vulnerability’ is strongly associated with inherent weaknesses of assets, whereas in engineering this association is largely absent. The next chapter shows that disciplines not only use different definitions but have also developed their own tools for understanding and assessing risk.

For this research, three disciplines in particular are relevant. First, there is the discipline of crisis and disaster management, the primary objective of crisis organisations on which this research is focused. For the purpose of examining terminology, tools, and methods, public health and safety and environmental protection are also included in this discipline. For telecommunication services two further disciplines are relevant. There is the discipline of engineering of large technical systems, and of large telecommunication infrastructures in particular. Business continuity is also included in this discipline. This is the second discipline in our list. It has a long history and gave rise to many tools that are still in use today. However, over the past decades a third discipline has risen in importance: that of information technology. Information technology has become so intertwined with telecommunications that nowadays the two are almost inseparable. The term Information and Communication Technology (ICT) covers this combined discipline.

It is therefore useful to look at definitions as used in the three disciplines of crisis management, engineering, and information technology. In order to be able to do so, the sources for these definitions are listed first.

2.1.1.1 Sources of terminology

This paragraph lists and describes, for each of the three disciplines, the sources of terms and definitions on which our choices for terminology are based. Completeness is not a goal here. Since there are so many definitions in use, it is sufficient to choose a few representative sources from each discipline, and to select from those for each term a definition that can be used consistently in the remainder of this thesis.

The United Nations International Strategy for Disaster Reduction (UN-ISDR) aims “at building disaster resilient communities by promoting increased awareness of the importance of disaster reduction as an integral component of sustainable development, with the goal of reducing human, social, economic and environmental losses due to natural hazards and related technological and environmental disasters”. One of its publications is the Terminology on Disaster Risk Reduction [159]. The United Nations University Institute for Environment and Human Security (UNU-EHS) has published a comparative glossary on risk, vulnerability and related concepts [156]. The World Health Organisation runs the International Programme on Chemical Safety (IPCS); the

<i>Crisis management</i>	<i>Engineering</i>	<i>Information Technology</i>
UN-ISDR Terminology on Disaster Risk Reduction [159]	ISO risk standards [85, 90]	Common criteria [77, 80]
UNU-EHS Comparative Glossary [156]	Society for Risk Analysis glossary [145]	ENISA compendium [39]
EEA online glossary [38]		Boehm [15]
IPCS Harmonization project [168]		Avižienis et al. [8]
IUPAC glossary in toxicology [35]		Leveson [107, 108]

Figure 2.1 – Sources of terminology from crisis management, engineering, and information technology.

IPCS Harmonization project has compiled an authoritative list of terms used in chemical hazards and risk assessment [168]. The European Environment Agency (EEA), an agency of the European Union, aims “to provide objective, reliable and comparable information, and the necessary technical and scientific support to the European Community and the member states” [122]. The EEA has compiled an extensive online database of environmental terminology [38]. Also, a comprehensive glossary of terms used in toxicology has been published by the International Union of Pure and Applied Chemistry (Duffus et al. [35]).

The International Organization for Standardization (ISO) has published several general risk-related standards for use in business and engineering [85, 90]. The Society for Risk Analysis publishes the well-known international academic journal *Risk Analysis*. It also maintains an extensive online glossary of risk-related terms [145].

The ISO has also published several standards specifically on security of information systems, e.g. the so-called Common Criteria [77] and the information security management systems series [80]. The European Network and Information Security Agency (ENISA), an agency of the European Union, is also active in this area. ENISA was created to ensure “a high and effective level of network and information security within the Community and in order to develop a culture of network and information security for the benefit of the citizens, consumers, enterprises and public sector organisations of the European Union, thus contributing to the smooth functioning of the internal market” [121]. ENISA has published a compendium of risk management principles, methods, and tools [39]. More in the discipline of information systems, risk in software engineering has been pioneered by Boehm [14, 15]. Terminology can be found in the taxonomy by Avižienis et al. [8]. Also, Leveson has published on safety of information systems [107, 108].

Figure 2.1 gives an overview. These sources will be perused while investigating terms and definitions in the following paragraphs.

2.1.2 Risk target, environment

First, the scope of risk and crisis management needs to be defined. What is the “thing” that needs protection? In the health discipline ‘organism’ is used for a single living being and ‘system’ for a collection of organisms [35, 38, 168]. Other documents use the term ‘system’ for the objects and processes that need protection, e.g. [156]. Very commonly, the term ‘environment’ is then used for anything that is not part of the system. (Care should be taken, because informal texts on ecology often use the term ‘the environment’ where this thesis would use ‘system’). In documents on engineering and (the security of) information systems, the terms ‘Target of Evaluation’, ‘Target of Assessment’, ‘IT system’, or ‘information system’ are often used [76, 77, 84]. System is then defined as “a specific IT installation, with a particular purpose and operational environment” [77].

risk target Of these definitions, the neutral *risk target* seems to best fit the needs of this thesis; ‘target’ and ‘target of assessment’ are also used. The term *environment* is used for anything that is not part of the target.

To understand and analyse the risk target, it is often useful to create a model. Leveson describes a model as “a representation of a system that can be manipulated in order to obtain information about the system itself” [107]. In this thesis, a model is description of the risk target that omits certain properties that are deemed irrelevant for analysis, but retains those properties that are relevant. Similarity between the model and the risk target allow statements about the risks in the model to be translated to risks to the risk target, through reasoning by analogy [46, 165].

2.1.3 Asset, stakeholder

asset An asset is described in general as “anything that has value to the organisation” [76]. In the context of information systems security, it has been described as “information or resources to be protected” [77]. In this thesis, an *asset* is an entity within the risk target that has value.

Common theme in these definitions, and implicit in other documents about asset protection, is the notion that assets represent a value, and that the value is not to be taken for granted.

Since owners value their assets and wish them to be protected, it is ownership that defines the boundaries of the system. From the ISO Common Criteria: “Safeguarding assets of interest is the responsibility of owners who place value on those assets” [77]. For human-made assets the matter of ownership is often clear, but for natural systems ownership is not always an appropriate concept. The term ‘responsibility’ may then be more general. It is, however, not always straight-forward to establish responsibility (or ownership) for an asset. The boundary between system and environment is often fuzzy, since systems interact with their environment. The interfaces are shared, and belong partly to the

system, and partly to the environment. A good illustration of this is the trend for organisations to network, creating interdependencies between organisations. Instead of a homogeneous environment an organisation now knows different gradations of “outsideness”, with some partner organisations being considered closer (less external) than others. Systems are also often capable of triggering external events through their interaction with the environment (e.g. leaking of confidential company information, causing the share price to plummet; or natural disasters partly caused by human actions). Under these conditions it becomes very difficult to say where responsibility for the system ends.

An important aspect of the risk target is its collection of stakeholders. A *stakeholder* is any person or organisation that places a legitimate value (by law or custom) on one or more assets. The owner of the risk target is an obvious stakeholder, as are the people or organisations responsible for protection of assets. Stakeholders are a diverse group. They can include, for example, concerned citizens in the neighbourhood of a factory. The value of an asset is an attribute of the relation between the asset and a stakeholder, and is therefore always subjective.

Note that assets can be tangible (such as land, buildings, or telecommunications equipment) or intangible (such as information, reputation, or biodiversity). The value of assets does not have to be expressed in financial terms. For example, according to ISO standard 27035 an incident response team “contributes to the reduction in physical and monetary damage, as well as the reduction of the damage to the organization’s reputation that is sometimes associated with information security incidents” [84].

The terms asset and stakeholder are adopted with the above meanings.

2.1.4 Hazard, safety

In engineering and information systems, threats are “the potential for abuse of protected assets” [77], “a potential cause of an unwanted incident which may result in harm to a system or organisation” [76], or a “risk source” [145]. Although not explicit in these definitions, there is often an assumption that threats are the result of deliberate human action. This assumption is made explicit in the ISO Common Criteria, which state that “The CC concentrates on threats to that information arising from human activities, whether malicious or otherwise, but may be applicable to some non-human threats as well”.

In the crisis management discipline threats are less often caused by deliberate human action, and the term ‘hazard’ is more common. A hazard is a “set of inherent properties of a substance, mixture of substances or a process involving substances that, under production, usage or disposal conditions, make it capable of causing adverse effects to organisms or the environment” [35], a “dangerous phenomenon, substance, human activity or condition that may cause loss of life, injury or other health impacts, property damage, loss of livelihoods and services,

social and economic disruption, or environmental damage” [159], an “inherent property of an agent or situation having the potential to cause adverse effects when an organism, system, or (sub)population is exposed to that agent” [168], “a threatening event, or the probability of occurrence of a potentially damaging phenomenon within a given time period and area” [38], or “a risk source where the potential consequences relate to [. . .] physical or psychological injury or damage”. Threats and hazards are thus largely synonymous, and a preference for one or the other depends mostly on whether the threat agent is human or natural, and whether the agent’s actions are deliberate or not.

Absence of natural threats (hazards) is usually described as safety, whereas absence of wilful human threats is called security. Duffus et al. [35] define (chemical) safety as “practical certainty that there will be no exposure of organisms to toxic amounts of any substance or group of substances”. The World Health Organization [168] defines safety as “practical certainty that adverse effects will not result from exposure to an agent under defined circumstances”. The SRA glossary defines ‘safe’ as “without unacceptable risks” and notes that this is sometimes limited to risks related to non-intentional events; ‘security’ is defined as “without unacceptable risks when restricting the concept of risk to intentional acts by intelligent actors” [145].

In safety circles the distinction is often made between internal safety and external safety. External safety is reached when hazards cannot affect the environment (e.g. people living close to a chemical plant).

Leveson [108] has argued that safety and reliability are different concepts. A system can be unreliable but safe, or unsafe but reliable. Increasing reliability will not, in itself, lead to increased safety.

In the research domain of this thesis, natural, inadvertent and unintended consequences are most common. The terms ‘hazard’ and ‘safety’ are therefore preferred, but without excluding the possibility of intentional attacks; the term *hazard* ‘threat’ is not used. *Hazard* is defined as “a potential cause of an unwanted incident which may result in harm to the risk target”, and *safety* as “the absence of hazards”. This thesis does not commonly use the term ‘security’.

Cause and effect chains

Hazards can have a cause that in itself can be considered a hazard. In general there can be a long chain of events leading up to a hazard. For example, a power failure (a hazard) can be caused by flooding of the area where electrical equipment is located; the flooding can be caused by a dike failure, caused by high waters, caused by a heavy storm during spring tide. Which of these events constitutes the main hazard depends on your perspective, which stems from your responsibilities (ownership) and thus your target boundaries. For a telecom engineer the main hazard will be the power failure. For the owner of the building the main hazard will be the flooding; the power failure has no

effects on assets within the owner's responsibilities, and therefore does not constitute a hazard. For government the hazard will be the breaching of the dike. Government may recognise the flooding of the building and the power failure as a hazard, depending on whether the building and the telecom facility have critical functions.

Similar to the chain of events leading up to a hazard, it is possible to identify a chain of events as a result of the loss. By definition, there always is a direct loss, and there may be secondary losses within the risk target or in its environment. In principle, the owner of a system is interested in the cumulative size of the entire chain of losses. In practice, owners choose to limit their analysis at a certain point in the effect chain, often the system boundary.

The next chapter (Section 3.1.1) shows that in complex systems the notion of a chain of events is insufficient, and it is more effective to view accidents as a result of system properties.

Positive and negative effects

Although the term hazard has negative connotations, hazards (and threats) can cause positive effects in addition to losses. The values of assets are subjective, and depend on the stakeholder. For example, effects on the security of a system that are considered positive by a human threat agent will be experienced as negative by the owner of the affected assets. For some stakeholders a hazards may be more accurately described as an 'opportunity'.

But also from the point of view of a single stakeholder there can be both positive and negative effects. A risky change in business plan may lead to many new opportunities, but might have adverse financial consequences. Adoption of a new oil drilling technique will save money, but may lead to environmental damage. This balance between positive and negative effects is one of the issues in risk evaluation (see Section 2.4.3). Note that it is possible for the immediate (primary) effect to be mostly negative, while secondary effects are mostly positive. The direct effects of a major earthquake will be disastrous, but afterwards there will be positive effects for the construction industry and for urban renewal.

2.1.5 Failure, Vulnerability

The terms 'failure' and 'error' are not defined nor used in the ISO risk standards [85, 90], but are common in literature on dependability and reliability. This thesis adopts Leveson's definition of *failure* as "the nonperformance or inability *failure* of the system or component to perform its intended function for a specified time under specified environmental conditions" [107]. Avižienis et al. define it as "an event that occurs when the delivered service deviates from correct service" [8]. A failure is therefore an event in which the system or component does not function as designed or planned. This deviation itself is called 'error'

common
cause failure

in Avižienis' terms; Leveson defines 'error' as "a design flaw or deviation from a desired service". The term 'error' is not used in this thesis. A fault is "the adjudged or hypothesized cause of an error" (Avižienis). *Common cause failures* are defined as "multiple component failures having the same cause" (Leveson).

The term 'vulnerability' connects hazards (or threats) to assets. Not all assets are equally affected by all hazards. For example, underground cables are easily damaged by trenching, but are not usually damaged by flooding of the surface area above them. The extent to which an asset can be affected by a hazard is called its vulnerability to that hazard.

In security of information systems and telecommunications, the ISO standards on information technology security techniques describe 'vulnerability' as [76]: "Vulnerabilities associated with assets include weaknesses in physical layout, organization, procedures, personnel, management, administration, hardware, software or information. They may be exploited by a threat that may cause harm to the IT system or business objectives. A vulnerability in itself does not cause harm; a vulnerability is merely a condition or set of conditions that may allow a threat to affect an asset." In engineering, 'vulnerability' is defined as "The characteristics and circumstances of a community, system or asset that make it susceptible to the damaging effects of a hazard" [159]. Thywissen [156] collected 36 descriptions of 'vulnerability'. Some of these include the number of options available to a community to deal with hazards. Vulnerable groups are those with the smallest coping ability who are therefore expected to suffer most when hazards materialise. It is notable that almost none of the definitions mentioned by Thywissen mention inherent flaws in assets as a contributing factor. In the discipline of health, 'vulnerability' is not commonly used at all.

vulner-
ability

We choose to define threats and hazards as external influences (from the point of view of the target), and vulnerabilities as properties of internal assets. A *vulnerability* in this thesis is defined as "a weakness of a component that, in combination with a hazard and under certain adverse conditions (e.g. a sufficiently motivated attacker, bad weather) will lead to failure".

Sometimes when the weakness and component are obvious, especially when discussing components in telecommunication service diagrams, we use the term vulnerability to denote just the hazard, e.g.: "power failure is one of the vulnerabilities of Router 2".

2.2 Uncertainty

Uncertainty is an essential ingredient of risk. If it is certain that a hazard will materialise, then we speak of fact rather than risk. If it is certain that a hazard will not materialise, then we need not spend time and effort on it. The likelihood of a hazard materialising is not the only uncertain aspect of risk. There is often considerable uncertainty about the magnitude of the vulnerability and effect,

and even uncertainty about the existence of assets, hazards, etc.

It is commonly agreed among risk analysts (e.g. [5, 67, 127, 145]) that two main types of uncertainty exist. First, there is the uncertainty when it is known that one correct value exists for a parameter, but the value cannot be determined because of practical limitations. For example, we know for certain that the world's population, at any one time, must be an integer number close to 7.3 billion. For obvious practical reasons, discovering the exact number at any point in time is difficult. This type of uncertainty arises from limited knowledge, even though in principle perfect knowledge is attainable, given unlimited resources. This type of uncertainty is therefore called *epistemic uncertainty*. Other terms used are subjective uncertainty, ignorance, and reducible uncertainty.

Second, there is the uncertainty when there is no single correct value, because the parameter is constantly changing (depending on the time or place of its assessment). Even when resources for measurement are plenty, every measurement would yield a different value. Dice rolls are the classical example; there is no single correct value for the roll of a die. This type of uncertainty arises from inherent natural variations, and is called *aleatory uncertainty*. Other terms used are stochastic uncertainty, probabilistic uncertainty, objective uncertainty, variability, and irreducible uncertainty.

Recently, researchers started to recognise *ambiguity* as a separate type of uncertainty [97, 160]. Ambiguity arises from differences in values held by stakeholders, which lead to different interpretations and perceptions.

A variable may be subject to all three types of uncertainty. The world's population figure, for example, is also variable; people pass away and babies are born every minute.

A closer examination of uncertainty may yield other types of uncertainties, although it is debatable whether these uncertainties should be classified in their own right, or whether they are subcategories of epistemic or aleatory uncertainty or ambiguity. For example, Cauvin et al. [20] consider the uncertainties arising from *modelling*. Any model is, by definition, an approximation of reality. There may be uncertainty on the correctness of the model. The results predicted by a model typically differ from the observed results. It is not known beforehand how big that difference will be, and therefore there exists uncertainty about the predicted results.

Another type of uncertainty arises from *measurements*. Any measurement has some inaccuracy, and there will be uncertainty on the true value.

Other authors (e.g. Colyvan [23]) consider uncertainties arising from the use of *natural language*. For example, consider the question: "What is the risk that our camping trip will be spoiled by rain?". The term "rain" is vague, as rain manifests itself in varying degrees, ranging from an occasional slight drizzle to a continuous downpour. It is possible to do away with vagueness by defining clear (and somewhat arbitrary) boundaries between "drizzle", "shower" and

“deluge”, but that will lead to edge conditions whereby a single drop triggers the change from “shower” to “deluge”. In addition to vagueness, natural language may suffer from under-specificity, and context dependence [23]. Virtually any expression in natural language can be a source of such uncertainty. Also note that different stakeholders will value the term “spoiled” differently; this an example of ambiguity.

Uncertainty can be modelled using various methods. Modelling of uncertainty will be addressed further in Section 2.4.2.

2.3 Risk

As illustrated at the start of this chapter, definitions of risk vary greatly.

In security of information systems and telecommunications, it is defined as “the potential that a given threat will exploit vulnerabilities of an asset or group of assets to cause loss or damage to the assets” [39, 76]. Boehm uses the term risk exposure, and defines it as the product of the probability of an unsatisfactory outcome and the loss incurred by it [15]. Risk exposure is therefore the expected value of the loss, in the statistical meaning of that term. UN-ISDR defines risk as “The combination of the probability of an event and its negative consequences” [159]. All these definitions combine a measure of uncertainty with the potential impact.

The ISO defines risk more generally as “the effect of uncertainty on objectives” in its Risk Management Vocabulary of 2009 [90]. This is an update of the definition in the 2002 version of the Vocabulary, where it defined risk as the “combination of the probability of an event and its consequence.” The new Vocabulary was published simultaneously with the first standard in an ISO series on risk management [85], which is based in large part on an Australian/New Zealand standard [147]. The AS/NZS 4360 standard was the first generic standard on risk management; it defined risk as “the chance of something happening that will have an impact on objectives”.

In the health and environmental discipline, Duffus et al. [35] give two definitions of risk: “Possibility that a harmful event (death, injury or loss) arising from exposure to a chemical or physical agent may occur under specific conditions” and “Expected frequency of occurrence of a harmful event (death, injury or loss) arising from exposure to a chemical or physical agent under specific conditions”. The World Health Organization [168] defines it as “The probability of an adverse effect in an organism, system, or (sub)population caused under specified circumstances by exposure to an agent”. The European Commission defines risk, in the context of health, as “the probability that an event will occur” [37]. Note that these definitions emphasise the uncertainty (the probability) and less so the expected impact of the hazard. This is generally the case within the health sciences [21].

<i>term</i>	<i>definition</i>
(risk) target	all assets that are within scope of a risk assessment
environment	anything that is not part of the risk target
asset	an entity within the risk target that has value
stakeholder	a person or organisation that places a legitimate value on one or more assets
hazard	a potential cause of an unwanted incident which may result in harm to the risk target
safety	the absence of hazards
failure	the nonperformance or inability of the system or component to perform its intended function for a specified time under specified environmental conditions
common cause failure	multiple component failures having the same cause
vulnerability	a weakness of a component that, in combination with a hazard and under certain adverse conditions (e.g. a sufficiently motivated attacker, bad weather) will lead to failure
risk	a set of triplets, each describing a scenario, its probability and its effect

Figure 2.2 – Definitions introduced in this chapter, and used in this thesis.

In this thesis we use a definition of *risk* that was given by Kaplan and Garrick [94]. They define risk as a set of triplets; each triplet describes a scenario (“What can happen?”), its probability (“How likely is it that that will happen?”) and its effect (“If it does happen, what are the consequences?”). This definition is a more detailed restatement of the ISO definition.

2.3.1 Facts and values

Distinction is often made between two views on risk: that of risk as a socially constructed concept (constructivism or social constructionism), versus that of risk as an objective attribute of physical entities (objectivism) [17, 59, 126].

Objectivism holds that risk is a consequence of the properties of objects and technology, and can be assessed objectively. In other words: risk is real. Independent assessors would, given the same information, arrive at the same assessment of risk. In this view, personal values have no place in risk assessment; risk assessment should be based on observable facts only [4, 100]. Personal values will certainly affect preferences for risk treatment options, but have no place in assessing the risk itself.

Constructivism holds that risk has no independent existence and is inextricably connected to the risk assessor. People create risks, by shaping their values and views around observations, and by interactions with others. In other words: risk is not real in itself, but receives reality and meaning within a social context. In this view, any attempt at objective risk assessment will be hampered by the assessor’s subjective choices and assumptions [143].

These two views are extremes. Practitioners most commonly take a flexible stance between constructivism and objectivism. It is advisable for public risk assessments to be as objective as possible, and to present choices and assumptions separately from the risk analysis. On the other hand, risk is in essence uncertain and assessment cannot be entirely based on hard facts so that implicit assumptions based on personal values are almost inevitable.

The two views on risk, objectivism and constructivism, are important to this thesis; the topic will be explored further in Chapter 4.

2.3.2 Risk perception and communication

When the potential effects of risk affect the public or society at large, other aspects of risk become relevant.

Perceived risk is risk as informally assessed, often by a lay person. Risk perception can be strongly influenced by, for example, fear or recent media exposure [28, 29, 47, 95, 137]. Although perceived risk is often portrayed as irrational or imaginary, its consequences may well be real. The so-called Thomas theorem (“If men define situations as real, they are real in their consequences” [155]) applies to perceived risk as well.

Perceived risk may exist objectively or may be socially constructed. A telecommunication-related illustration of this is the ongoing discussion on health effects of electromagnetic fields (this topic will return in Section 4.2.1). Objectivists hold that ample scientific research failed to show any evidence of health effects, and that the risk — if it exists at all — is negligible. Despite this research, others maintain, often in emotive language, that a large amount of anecdotal evidence warrants precaution and lowering of permissible emission levels.

Risk communication is the dissemination of information about certain risks to the general public. Risk communication helps to raise awareness about those risks, and increases resilience of communities. Trust in authorities is an essential ingredient to successful risk communication [30, 104, 144].

Risk communication should not be confused with crisis communication. Crisis communication advises the general public and other stakeholders about actions to take in an occurring crisis. Its goal is to assist in crisis response, whereas risk communication aims to improve pro-action, prevention, and preparation.

2.4 Risk management

Risk management can loosely be described as a management activity for mitigation of hazards and threats. For terminology related to risk management, the ISO standards are used again.

The ISO series 31000 is an international generic standard for risk management

[85, 86]. It is based on an earlier Australian/New Zealand standard on risk management [147]. It describes the principles, a management framework (according to the plan–do–check–act cycle of continuous improvement), and the outline of a risk management process. The core process consists of the following steps (also see *Figure 2.3*).

1. Establishing the context. By establishing the context, the organisation articulates its objectives, defines the external and internal parameters to be taken into account when managing risk, and sets the scope and risk criteria for the remaining process.
2. Risk identification. The aim of this step is to generate a comprehensive list of risks based on those events that might create, prevent, enhance, degrade, accelerate or delay the achievement of objectives. It is important to also identify the risks associated with not pursuing an opportunity.
3. Risk analysis. Risk analysis involves consideration of the causes and sources of risk, their positive and negative consequences, the likelihood that those consequences can occur, and other factors that influence risk evaluation.
4. Risk evaluation. Risk evaluation involves comparing the level of risk found during the analysis process with risk criteria established when the context was considered. Based on this comparison, the need for treatment can be considered.
5. Risk treatment. Risk treatment involves selecting one or more options for modifying risks, and implementing those options.

These steps are continuously complemented by the supporting activities communication and consultation, and monitoring and review. Risk identification, analysis and evaluation are collectively called risk assessment. Risk treatments that deal with negative consequences (as opposed to simply accepting them) are also referred to as risk mitigation, risk elimination, risk prevention and risk reduction.

Boehm's work on software engineering risk management uses a similar approach with some different terminology (e.g. "risk prioritisation" instead of "risk evaluation") [15]. The main difference is that his approach is not continuous but dynamic; instead of redoing the risk assessment after implementing risk treatments, Boehm proposes to perform risk assessment on the *changes*. This works well for risk-driven software-process models, where the development methodology is based on risk management [14, 125]. In the operational phase of systems, when the risk target is relatively stable, the iterative approach by ISO is common.

Figure 2.3 – The risk management process in the ISO standards on Risk Management [85].

2.4.1 Risk identification

During risk identification, the risk professionals enumerate all the risks that could affect the risk target under their responsibility. Risk discovery includes listing all possible hazards, threats, assets, and areas of consequences. Comprehensive identification is critical, because a risk that is not identified at this stage will not be included in further analysis. In practice, it often happens that a risk is excluded from analysis because it is deemed to be impossible or extremely rare, but that it then unexpectedly happens. Uncertainty therefore makes risk identification difficult. When uncertainty is of the ambiguity type, risk identification becomes a social as well as a physical question [97].

Risk identification techniques include Failure Mode and Effect Analysis (FMEA) [87], Hazard and Operability Studies (HAZOP) [89], Preliminary Hazard Analysis (PHA), Human Reliability Analysis, brainstorming, structured interviews, checklists, or scenario analysis. Some of these are examined in the next chapter.

2.4.2 Risk analysis

In the previous stage risk professionals created a list of risks to include in their analysis. During risk analysis, all available knowledge about these risks is compiled. This may include description of hazards, chains of causes and chain of effects, and quantification of the vulnerabilities in the risk target and all possible effects. Risk analysis provides the necessary knowledge to risk evaluation and decisions on risk treatment.

Risk analysis techniques include FMEA, Fault Tree Analysis (FTA) [88], HAZOP (effects only), Event Tree Analysis, and Monte Carlo simulation techniques. Some of these are examined in the next chapter.

Role of experts

Risk analysis is a specialist activity. It requires an in depth understanding of the risk target and of the complex internals of its workings. Experts can list hazards, indicate vulnerabilities, estimate parameters, and explain the workings of the risk target in general. Experts can also be asked to estimate their own uncertainty on these answers. Since the experts who are involved in risk analysis are typically not the persons responsible for risk treatment, communicating the knowledge on risk is therefore important. Analysts must take care that their level of confidence and the impact of uncertainties is made clear and understood.

Expert assessments can be biased. Cognitive biases have been studied extensively, most notably by Tversky and Kahneman [93, 157]. Well-known examples are confirmation bias (when new information confirming existing opinions is valued higher than new information contradicting existing opinions) and “group think” (when a person derives his or her opinions from those of the social group). People are usually bad at estimating their own biases.

Peer reviews have been suggested as an essential tool when involving experts [4]. There are also procedures for expert judgement elicitation that compensate for biases by weighting estimates from different experts based on their past performance [24, 25].

Chapter 7 investigates the reliability of expert estimates in RASTER.

Modelling of uncertainties

An important issue is how uncertainties are modelled. Different authors choose different approaches, often depending on the type of uncertainty.

It is widely accepted that aleatory uncertainties can accurately be modelled using probability theory. Uncertain parameters are then estimated using a probability distribution function. For example, a single die roll can be modelled with a uniform probability distribution.

Several methods have been proposed to deal with epistemic uncertainty [63]. Probability theory is considered adequate by several authors (e.g. [7]), although others disagree (e.g. [42]). This controversy remains unresolved. Several alternatives have been put forwards, such as interval analysis and possibility theory. These will not be discussed further.

For uncertainties in natural language fuzzy logic has been used to the satisfaction of several authors (e.g. [23]).

Meaning of probability

It should also be noted that there exist different schools of thought on what probability actually means. Two main views are the frequentist interpretation and the subjective interpretation.

The *frequentist interpretation* is that probabilities reflect the relative incidence of events. If it has been observed that a six is rolled in 15% of all die rolls, then

that defines the probability of rolling a six. Frequentist probabilities derive from a body of observations, and is therefore the prevalent interpretation in statistics. A frequentist probability is a statement about the world, about an inherent property of the objects or the process; frequentist probabilities are therefore also called objective probabilities. Strictly speaking, the probabilities of a single event or of future events are meaningless in this view, as these cannot be expressed as a relative frequency of observations. In a somewhat relaxed interpretation, the body of observations may be a hypothetical set of infinitely repeated independent processes. If a fair die would be thrown an infinite number of times, the probability of rolling a six would get ever closer to $\frac{1}{6}$. As the body of observations expands, epistemic uncertainty becomes smaller, and aleatory uncertainty can be more accurately described. Well-known frequentists are Fisher and Von Mises [50].

In the *subjective interpretation* probabilities reflect a personal degree of belief [32]. If I have a less than total confidence that it will rain tomorrow, I can assign the probability 90% to that event. Probabilities are based on available information, and as more information is collected the subjective probability will be adjusted. With the subjective view, different observers may assign different probabilities to the same proposition as each observer possesses different information. Subjective probabilities are a statement about the credibility of the evidence to support a statement, about the confidence in the statement; they are a property of the subject. There is no uncertainty about subjective probabilities, but there may be disagreement among subjects. Well-known 'subjectivists' are Bayes, De Finetti, and Savage [50].

Both the frequentist and the subjective interpretation of a rational observer are compatible with axiomatic definitions of probability by Kolmogorov. The frequentist and subjective interpretations are also not at odds with each other: as more information becomes available and the body of observations increases, the frequentist and subjective probabilities will converge [55].

Like the two different views on risk (objectivism and constructivism), the two interpretations of probability are important to this thesis. These topics will be discussed in Chapter 4.

2.4.3 Risk evaluation

Based on the available knowledge about risks compiled during risk analysis, risk evaluation determines the magnitude of the risk and compares it to the acceptable level and expected benefits. If quantification is not possible, other methods such as rankings and expert consultations may be used to obtain estimates of risk levels.

The goal of risk evaluation is to decide which risks are important enough to warrant risk treatment. The decision must also take into account the legal requirements, regulations, internal policies, moral obligations, social responsi-

<i>Risk treatment options</i>	
avoidance	deciding not to start or continue with the activity that gives rise to the risk, <i>or</i> eliminating the risk source.
reduction	changing the likelihood, <i>or</i> changing the consequences.
transfer	sharing the risk with another party or parties (e.g. through insurance).
retention	accepting (“taking”) or increasing the risk, in order to pursue an opportunity, <i>or</i> retain it by informed decision.

Figure 2.4 – The four (left column) or seven (right column) risk treatment options commonly recognised.

bility, and the interests of other stakeholders. Subjectivity and uncertainty may make this a difficult job. To reduce bias, the persons doing the risk evaluation should avoid being involved in the risk mitigation as well; their involvement could easily lead to (suspicions of) wishful thinking, political interference, or other influences (such as the inclusion of extraneous commercial interests).

Techniques for risk evaluation include FMEA and Monte Carlo simulations.

2.4.4 Risk treatment

Risk evaluation decided on those risks that require treatment. During the risk treatment stage suitable countermeasures are identified, chosen, and implemented. The ISO Risk Management standard lists seven types of countermeasures which can be used in isolation or in combination. Other authors and ISO standards (e.g. [5, 76]) use a more concise list of four types. These four types, and the corresponding seven types from the ISO standard are shown in Figure 2.4.

Since risk treatment may modify the risk target, it could introduce new risks. The environment of the risk target may change over time. Risk management should therefore be a continuous process.

2.5 Conclusion

Terminology in this research domain lacks a single, authoritative source of definitions. Definitions differ among academic disciplines, and even within one discipline sometimes multiple definitions are in use. This chapter therefore examined risk-related terminology from three disciplines that are relevant to our problem domain: crisis management, engineering, and information technology. From these, definitions were selected to be used in a consistent manner in the

remainder of this thesis: see *Figure 2.2*.

The concepts of uncertainty, risk, and risk management were examined and discussed. Of particular interest are two views on risk: objectivism and constructivism. Also, there are two views of the meaning of probability: the frequentist and the subjective interpretation. These views and interpretations will be discussed further in Chapter 4.

Current Practice and Theory

A summary of the first section of this chapter appeared in the author's publication "Design and initial validation of the Raster method for telecom service availability risk assessment" in the proceedings of the 9th International Conference on Information Systems for Crisis Response and Management ISCRAM 2012.

Research question RQ1 concerns current risk assessment methods. The topics of risk assessment and risk management have a long history. Several tools and methods have been developed over the years for understanding and assessing risks. This chapter gives an overview of the methods currently in use: first as a general overview, then, in the second part, by investigating a few specific methods in detail.

3.1 Overview

This section gives a general overview of risk assessment and risk management methods. For this, the three disciplines are used that also guided the investigations on terminology in the previous chapter: those of crisis management, engineering, and information technology. This section follows a rough chronological order: first engineering, then crisis management, and finally information technology, being the youngest discipline of the three.

3.1.1 Engineering

Safety of technical and industrial systems and of their operators (then also called occupational health and safety) became an issue at the start of the industrial revolution. New technology at that time led to novel kinds of accidents, which indicated the emergence of new risks. Risk assessment was recognised as a necessary activity to prevent losses. Hale and Hovden describe this as the

start of the *first age of safety* [56]. The remainder of this paragraph follows their description of the three ages of safety.

Risk assessment in the first age (see *Figure 3.1* for its approximate period) was characterised by a focus on the technical aspects of systems. Accident models treated risk as a linear propagation of failures, as a chain of events; for example, Heinrich's domino model (1931, [72]). Many risk assessment methods were developed in this age, including Failure Mode and Effects Analysis (FMEA, 1949, [87]), Hazard and operability studies (HAZOP, 1960s, [89]), Fault Tree Analysis (FTA, 1962, [88]), Event Tree Analysis (1970, [135]), and Management Oversight and Risk Tree (MORT, 1973, [91]). Iconic application areas for this age were nuclear energy and nuclear weapon systems, as risks in these areas potentially have a high impact.

The idea of performing a probabilistic safety assessment (sometimes as required by regulations) also arose from the first age. These assessments featured quantitative estimates of probability and effects, and used the statistical notion of expected loss as the measure of risk. This is similar to Boehm's approach to software engineering risks [15]. This approach is now sometimes called *classical risk analysis* [6, 127, 162].

These risk assessment methods are still in active use today, and the International Electrotechnical Commission actively maintains them as international standards. However, accidents proved that these methods were insufficient to mitigate non-classical risks. Iconic examples are the accident in the Seveso chemical plant (Italy, 1976), and the accident at the Three Mile Island nuclear power plant (1979). These accidents were not purely technical in nature, but prominently featured interaction between human operators and technical systems. Perrow typifies these accidents as system accidents or "normal accidents" (1984, [128]). Normal accidents according to Perrow are inevitable in systems that have both complex interactions and tight coupling between subsystems.

The *second age of safety*, in Hale and Hovden's typology, started when risk assessment methods began to include human failures and human reliability. Purely technical systems were now seen as human-technical systems, and human reliability assessment became a separate topic of study. Human reliability assessment treated human failures in the same way as failure of technical components, with a fixed failure probability for certain types of tasks. Just as a pipe was modelled with a certain failure rate depending on the corrosiveness and temperature of the fluid carried by it, so were human operators supposed to have a failure rate depending on their task. Methods include Technique for Human Error Rate Prediction (THERP, 1983, [150]) and Human Error Assessment and Reduction Technique (HEART, 1986, [166]). Reason's "swiss cheese model" (1990, [136]) is an example of an accident models that explicitly addresses human reliability.

Although human reliability assessment methods were useful, with time

<i>Age</i>	<i>Accident model</i>	<i>Examples</i>
First age: 1800 – 1980	Technical: accidents as a linear sequence of events. Limited to physical components.	FMEA, HAZOP, FTA, Event trees, domino model.
Second age: 1980 – 1995	Human interaction: accidents as a failure of interaction between human operator and technical system. First generation human reliability models: failure probability for each type of interaction. Limited to simple model of human behaviour.	THERP, HEART, “swiss cheese” model.
Third age: 1995 onwards	Organisation or system oriented: accidents as a failure within a complex socio-technical system. Second generation human reliability models: failures in context of external circumstances. Methods are complex and time consuming.	STAMP, FRAM, ATHEANA, MERMOS, CREAM.

Figure 3.1 – The three ages of safety, from Hale and Hovden [56]. See main text for meaning of acronyms and references. Dates are indicative.

criticism was raised at these methods. They became known as first generation methods after a new generation of methods emerged [34]. These second generation methods paid consideration to the interaction between environmental circumstances and human performance. Methods include A Technique for Human Error Analysis (ATHEANA, 1996, [26]), MERMOS (1997, [103]), and Cognitive Reliability Error Analysis Method (CREAM, 1998, [70]). In Hale and Hovden’s typology, these methods belong to the *third age of safety*: that of safety management and safety culture. Risk assessment in the third age takes on a systems view that extends beyond technical components and human operators, and takes organisational structures into account. LaPorte studied organisations that managed to avoid Perrow’s “normal accidents” by focussing on a strong safety culture (1991, [102]). He called these “high reliability organisations” (HROs). If the accidents at Seveso and Three Mile Island were iconic for the second age, then icons for the start of the third age of safety were the accidents with the space shuttles Challenger (1986) and Columbia (2003). New risk models emerged for the third age. In these risk models, risks are viewed as the result of non-linear interactions; safety is seen as an emergent property, a system property. This is a departure from the first and second age, in which accidents were assumed to have an attributable cause: technical in the case of the first age, or human in the case of the second age. A typical third-age risk assessment method is Leveson’s Systems-Theoretic Accident Model and Process STAMP (2003, [108, 110]). In STAMP, an accident is viewed as a failure of the system to maintain safety constraints. Another example is Hollnagel’s Functional Resonance Accident Model FRAM (2004, [69]). In FRAM, an accident is viewed as an unexpected peak or dip in normal fluctuations of system performance.

Each age of safety is therefore characterised by the scope of its accident models: technical for the first age, human factors for the second age, and system or organisational for the third age. Transitions between ages arose from an emerging understanding that the contemporary accident models were in some way insufficient. Such understanding was often triggered, at least in the public's mind, by iconic accidents that demonstrated the shortcomings in the state of the art. *Figure 3.1* summarises this development. It is no surprise that current literature speculates on the next developments: the fourth or fifth age of safety [16, 51, 134].

Organisations can apply general risk management standards; one of the most well-known is the IEC/ISO 31000 standard (2009, [85]). This standard builds upon earlier work by the Australian/New Zealand Standard AS/NZS 4360 (2004, [147]). Note that these standards are general approaches that require a lot of customisation before they can be implemented in an organisation. They are not step-by-step methods that can be readily applied to achieve an acceptable state of risk management. In addition to those general methods, there are also several industry-specific standards, for example those tailored to the information and communication technology industry.

3.1.2 Crisis management

The development of risk assessment methods has been strongly influenced by government regulations on safety. In the late 18th century the first labour regulations on restrictions of child labour arose in response to what were considered unacceptable working conditions. Initially regulations did not regard safety, but legislation to protect workers against physical harm appeared at the start of the 19th century.

During the 1960s, society in general became aware of the inadequacy of the state of the art in risk analysis and risk management. An iconic example is Carson's book *Silent Spring* on the effect of pesticides on ecology (1962, [19]). Risk management was no longer seen as an activity that could be left to industry and private litigation. Self-regulation proved to be insufficient to protect people, the environment, and future generations. This situation prevailed into the third age of safety. Accidents in chemical plants in Flixborough (1974), Seveso (1976) and Bhopal (1984) led to new legislation such as the Seveso and Seveso II directives (1982, 1996, [120]), and continued governmental studies (e.g. [36, 68]).

Risk governance takes a broader view on risk management that involves all stakeholders, including government and citizens, and takes into account social, economic, environmental and political issues in addition to commercial and financial interests. When risk governance is applied to new or emerging technologies the term 'technology assessment' is used. Examples of risk governance frameworks are the WBGU method (2002, [97]) and the PBRC scheme (2006, [100]).

One of the ideas stemming from risk governance is the *precautionary principle* [138, 146]. Informally, this principle states that even when risks are uncertain the possibility of irreversible harm should be avoided. The United Nations Rio Declaration stated it as “Where there are threats of serious or irreversible damage, lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures. . . ” (1992, [158]). The precautionary principle is also at the basis of the European environmental policy, embedded in Article 130r of the Maastricht Treaty (1992, [40]).

The German Advisory Council on Global Change (*Wissenschaftliche Beirat der Bundesregierung Globale Umweltveränderungen*, WBGU) proposed a risk evaluation and classification method that deals explicitly with social issues and public opinion (1998, [48]). Their method recognises uncertainty as not simply a complicating factor to risk assessment, but as a factor that gives direction to risk treatment. The precautionary principle is an important part of their proposed method, for certain kinds of risks. The Dutch Scientific Council for Government Policy (*Wetenschappelijke Raad voor het Regeringsbeleid*, WRR) later endorsed this concept in their in their advisory report to the Dutch government (2008, [36]).

In the Netherlands, Safety Regions have to make a risk profile for their region, but the law does not prescribe any particular method. This was intentionally left up to the management of each Safety Region. The Memorandum refers to “several manuals and guidelines”, but only explicitly refers to the Guidebook Reference Disasters (*Leidraad Maatramp*, 2000, [73]). This guidebook documents a method for disaster identification, assessing possible effects, and determining the maximum necessary response capacity. The method is therefore based on derived worst-case scenarios.

Safety Regions have a legal obligation to plan for unavailability of electrical power, information systems and telecommunications. A study commissioned by the Ministry of Safety and Justice revealed that in 2012 most Safety Regions did not have such a plan [33]. The ministry organised drafting sessions, based on a *continuity plan framework* to stimulate compliance and improve the quality of plans.

3.1.3 Information systems

One particular and new kind of technological system is created by information and communication technology (ICT). The rapid uptake of the internet has been particularly disruptive. For some time some people believed “cyberspace” to be entirely independent from the physical world; Barlow’s declaration of independence of cyberspace is an iconic example of this school of thought (1996, [10]). However, it quickly became apparent that accidents in cyberspace cause damage in the real world, an area where governments wield undeniable influence and power (2006, [53]). Nowadays the notion of cyberspace as an independent domain has all but disappeared, but nevertheless the internet is

undoubtedly one of the most significant inventions in recent history. Interest in internet governance is on the increase, for example through the Global Forum on Cyber Expertise and the Governmental Advisory Committee of ICANN [52].

The different nature of vulnerabilities of ICT systems, and the novel kinds of threats made new risk management methods necessary. Unlike industrial safety, ICT systems face deliberate attacks, from adversaries who exploit vulnerabilities. Those adversaries do not need to be physically present, but can literally be located at the other end of the world. And unlike industrial systems, the vulnerabilities are not physical but logical in nature; defects do not involve weak pipes or inadequate fencing but software flaws, compromised firewalls, and the like. Assets are not tangible resources or end-products, but intangible data of which confidentiality, integrity and availability have to be protected. New risk assessment and risk management methods were needed for this domain, and academic institutions and standardisation bodies rose to the task.

Some of the most well known standards are the ISO 27000 series on information security risk management (2005 onwards, [80]), and the ISO series on criteria for IT security evaluation, better known as the “common criteria” (1999 onwards, [77]). Other risk management methodologies are the CCTA Risk Analysis and Management Method CRAMM (1992, [9]), and the CORAS framework (2002, [45, 111]).

3.2 Selected methods and standards

The previous section gives an overview of techniques used in risk management, in historical perspective. This section investigates in more detail selected methods and standards that are especially relevant to our domain.

Methods that analyse possible events and their consequences can be divided into four basic types. First, there are *top-down* methods, in which system-level events are analysed by detailing their causes on lower levels in terms of part and component failures. *Bottom-up* methods trace an event from its lowest level effects up to system-level effects. Top-down and bottom-up methods view an event in a single point in time. There also exist methods that analyse events into consequences over time: methods based on *forward-logic*. Methods that analyse events from contributing causes backwards in time are called *backward-logic* methods. Of course combinations are also possible. Backward and forward logic can be applied to the same event, to investigate both its causes and consequences. Top-down and bottom-up methods can be augmented with backward or forward logic.

3.2.1 Engineering

From the first age of safety a few risk assessment methods are selected that are still in widespread use today; also one is selected from the third age of safety.

Failure Mode and Effect Analysis

The origins of Failure Mode Effect Analysis (FMEA) lie in reliability engineering. It was first used in the 1940s for military systems. In the 1960s it was successfully used in the Apollo Space Program, and in the 1970s its use was extended to automotive applications. It is now an international standard published by the International Electrotechnical Commission IEC [87].

FMEA can be used as soon as the design of the system is known, for example when the final design is available as a block diagram. At minimum the list of all system elements and the logical connections among them must be known. After preparation and startup, the analysis starts by listing all the ways in which the system could fail to perform its functions. Each “way in which to fail” is called a failure mode. For example, for a telephone exchange possible failure modes are “does not connect caller to called number”, “connects caller to wrong number”, or “premature end of call”. For each failure mode the immediate effects and ultimate effect is then listed. This gives an indication how serious the failure mode is. Then, the possible or most likely causes for the failure mode are listed. If a failure mode is serious enough, possible mitigating actions can optionally be suggested. This examination of effects and causes is performed for each identified failure mode. FMEA therefore is a forward-logic method, with top-down elements and backward-logic to estimate the risk severity.

Failure Mode, Effect and Criticality Analysis (FMECA) is an optional extension of FMEA in which values are assigned to the severity and detectability of effects and to the likelihood of causes. These values can be used to calculate risk priorities.

FMEA’s international standard mentions four limitations of the method [87]. First, FMEA assumes independent failures. Some provisions are made for common cause failures, but support is limited. Secondly, FMEA doesn’t gracefully handle human factors; as human operators are not considered part of the technical system, their “failure” to perform cannot easily be expressed. Thirdly, FMEA was developed with hardware systems in mind and does not readily lend itself to application to systems in which software plays an important role. Fourth, FMEA can be laborious and inefficient when applied to complex systems or when used indiscriminately. In addition to these limitations, it can be observed that FMEA does not provide a systematic approach to identify all failure modes, or causes, or effects. FMEA relies on experts for enumerating these, but the method itself provides little guidance. If, for example, a possible cause is unintentionally omitted the analysis may miss important risks and its outcome can be misleadingly positive.

Fault Tree Analysis

Fault Tree Analysis (FTA) was first used in 1961 for a missile launch control system. It has since seen widespread use in aerospace industry, and is nowadays

an international standard published by the IEC [88].

FTA can be used after design stage; the basic system components must be known. The analysts must first enumerate all the possible ways for the system to fail: the list of failure modes. For each failure modes in turn, FTA investigates its causes or combination of causes in terms of failure of lower-level components. Each of these causes is itself analysed as a failure mode. This decomposition ends when, at the lowest level, basic events are reached. FTA is therefore a top-down method. The decomposition can be modelled as a tree, in which nodes are causes. Sub-trees can either have an OR-relation (either of the sub-nodes can cause the failure mode) or an AND-relation (all of the sub-nodes must fail in order for the failure mode to happen). FTA can be used quantitatively, as it was originally designed, but can also be used qualitatively; the latter is more common.

FTA and FMEA can be used in combination; FMEA is then often used as a starting point for an FTA analysis.

FTA suffers from many of the same limitations as FMEA.

Hazard and Operability Studies

A limitation of FMEA and FTA is that they do not provide a method to discover the top-level events or failure modes. Hazard and Operability Analysis (HAZOP) is a method to identify risks. HAZOP originated in the 1960s, in safety analysis of chemical plants. It is now an international standard published by the IEC [89].

HAZOP can be applied on a system as soon as the design is complete. Qualitative guidewords such as *More* and *Early* are used to discover all possible deviations in the behaviour of system parts. By inspecting all system parts using each of the predefined guidewords, failure modes can be enumerated in a systematic creative process.

System-Theoretic Accident Model and Processes

Leveson has proposed System-theoretic Process Analysis (STPA) as a suitable risk assessment method for complex socio-technical systems [105, 108, 109]. STPA is an extension of Leveson's accident model STAMP (Systems-Theoretic Accident Model and Process). STAMP is based on the premise that accidents are caused by a failure of control processes to safely handle deviations from the normal flow of events. Such deviations can, for example, be technical component failures, unexpected software behaviour, or confusion because of miscommunication between human operators. In outline, STPA uses the following sequence of activities:

1. Establish fundamentals (define accidents, hazards, and system boundaries; draw high-level control structure).
2. Identify potentially unsafe control actions; create safety requirements and constraints.

3. Determine how each potentially hazardous control action could occur.

The control structure is not restricted to the technical components; organisational and societal control structures are included too. Unsafe control actions can therefore be automated controls, operator behaviour, as well as organisational processes and safety culture. In STPA there are four modes in which a control action can be unsafe:

- A required control action is not provided. For example, when driving a car: ignoring a red traffic light instead of stopping and waiting for the green light, or failure to obey a traffic sign.
- An unsafe control action is provided. For example: braking hard on a wet road, or overtaking against oncoming traffic.
- A required control action is provided too late, too early, or out of order. For example: using the indicators of a car while switching lanes (instead of in advance), or sequence mistakes in using the clutch and gear shift.
- A required control action is continued for too long, or stopped too early. For example, overfilling engine oil in response to a 'oil low' warning. When control actions are discrete rather than continuous this mode is not applicable.

Step 2 is concluded by the creation of safety requirements and constraints. Constraints are requirements on the system, stating that a certain (unsafe) state or action should never happen. The system design must ensure that these requirements are met, so that the potential hazards are fully mitigated and cannot occur. For example, the hazard of ignoring a red traffic light can be restated as the constraint "The car should be stationary whenever the traffic light is not green."

When system hazards have thus been identified, the control system is examined to see how the unsafe condition could arise. Unsafe control actions occur, by definition, in a control loop containing the controlled entity, the controller, actuators, and sensors. An unsafe control action can occur for several reasons, shown in *Figure 3.2*. For each unsafe control action the corresponding control loop is inspected for possible occurrence of any of the control failure causes. If none apply, the control action will always be safe and the hazard is in fact non-existent or fully mitigated. If at least one failure cause is possible, then this constitutes a risk. Mitigation actions can then be proposed, to remove the failure causes, or reduce their likelihood or impact.

Although STAMP has been applied successfully in large real-life accident investigations, the method is not easy to apply. Work has started on a primer, to assist analysts in applying the method to the risk target under their investigation [106].

Figure 3.2 – The STAMP control loop with all causal factors for an unsafe control action (from [109]).

3.2.2 Crisis management

Two methods are selected that are of special relevance to the Netherlands.

WBGU method for risk governance

There are two issues that make classical risk analysis less suitable for risk governance [36]. The first issue is that several assumptions that underly the classical approach are invalid for more complex risks. These assumptions are, for example, that risk assessment and risk treatment must be conducted separately, that experts possess the required knowledge for risk assessment, and that understanding risks means that they can be controlled and managed. When these assumptions do not hold, classical risk management as a tool is insufficient to manage and control risks. The second issue arises from fundamental changes in the global risk landscape. We live in a “risk society” [12]. Incidents are no longer seen as “acts of God”, but as a failure of responsibility. Industrial risks are no longer restricted to a small physical area and a short time span, but can be planet-wide and utterly devastating. Sometimes the consequences of our actions become apparent only much later, when it is (almost) too late to remove the risk source, e.g. depletion of the ozone layer due to CFCs, or human-induced climate change.

Cassandra: high impact and likelihood, but a large delayed effect.

Cyclops: well defined maximum damage, but largely uncertain probability.

Medusa: low classical risk, but frightening and with a high potential for mobilisation.

Pandora's Box: high ubiquity, persistency, and irreversibility.

Pythia: large uncertainty in both damage and probability.

Sword of Damocles: extreme impact.

Figure 3.3 – Risk classes in the WBGU method for risk governance [48].

The Explanatory Memorandum of the Dutch Safety Regions Act of 2010 [116] points to a shift from classical risks to more complex risks as one of the main reasons for introducing the new Act. Existing crisis organisation structures and regulations are not suited to risks that, for example, lack a clear relation between hazards and consequences, and are not clearly bounded in duration and location.

The approach proposed by the WBGU and endorsed by the WRR complements classical risk management so that complex, non-classical risks can be addressed as well. In the WBGU method, risk evaluation takes into account objective physical risk factors, risk perception and public concern. Based on these evaluation criteria, six risk types can be identified that are not well suited to the classical approach. These types were named after figures from Greek mythology (see Figure 3.3).

Risks can then be divided into four main risk classes, based on properties of

the damage, uncertainties, and on complexity of the decision making process.

- *Simple risks.* Simple, or “classical”, risks are characterised by well-known hazards and effects, likelihoods that are within common experience (neither extremely low nor high), and little disagreement on the best course of action. Damage can be repaired (it is reversible, and persistency and ubiquity are small), and there is little potential for social conflict. Experts’ estimates can be used to guide the risk mitigation, and well-known statistical methods can be used.
- *Complex risks.* Complex risks are characterised by high uncertainty about the cause–effect chain. Damages may be ubiquitous or persistent, or have a catastrophic potential. There may be a difference of opinion on treatment of the risk.
- *Uncertain risks.* Uncertain risks are characterised by very high uncertainty on the likelihood or the magnitude of effects, preventing any classical risk management.
- *Ambiguous risks.* Ambiguous risks are characterised by a lack of consensus. They are perceived in completely different ways by different stakeholders. Some stakeholders may think the risk to be negligible, while other stakeholders may believe the risks to be extremely high.

Each risk class then calls for an appropriate combination of risk mitigation strategies [5, 97].

The *risk based strategy* is based on the thought that all countermeasures carry a cost, and that a rational choice between benefits, the expected losses and the cost of countermeasures gives optimal overall results. This strategy is employed in classical risk management.

The *precautionary strategy* is based on the thought that the risk identification or analysis stages may fail to adequately predict the true risk level, or may even incorrectly state that no unacceptable risks exist. Uncertainty plays a leading role in the precautionary approach. The precautionary principle is applied, but complemented with continuous risk monitoring and further studies. The precautionary strategy is therefore an active strategy, not simply an avoiding of risks.

The *discursive strategy* is based on the thought that social aspects play a large role in all stages of risk management. Taking into account the opinions of society is considered essential in reaching a level of risk that is not only safe, but also perceived to be safe, and accepted. Open communication and discussion with all stakeholders is therefore essential in order to reach a risk level that is considered optimal by society.

The classical risk mitigation strategies are only suitable for classical risks. Complex risks are best addressed by classical risk mitigation strategies, but

Strategy	Simple risk	Complex risk	Uncertain risk	Ambiguous risk
risk-based	X	X		
precautionary			X	
discursive		science-based	stakeholders	all of society

Figure 3.4 – Risk mitigation strategies suitable for the different risk classes in the WBGU method.

complemented by science-based discursive strategies in which a diverse group of experts participates. For uncertain risks, the precautionary mitigation strategy is most suitable. Constant re-evaluation and a discourse in which all stakeholders participate are necessary. Ambiguous risks require mostly a discursive approach in which also the lay public participates, and a balancing of different risks. See Figure 3.4.

Continuity plan framework

One method is of particular interest, because it is a risk assessment method that was specifically designed to the needs of Dutch Safety Regions [13]. This method is based on a framework that assists in creating a continuity plan.

The method involves *critical processes, applications, and hardware components*. A critical process is one that supports a primary function of the Safety Region that must at all times be continued, such as basic fire services and the emergency call centre. Each critical process requires and depends on one or more applications. An application can either be an IT system or a telecommunications system. Each application requires and depends on one or more hardware components. Each of these components receives a three-level score to indicate its overall reliability:

- Level 1. Not resistant to component faults and unavailable during maintenance: components are not redundant and there are no double supply lines, hardware has no backup power (and is therefore not available during failures or maintenance).
- Level 2. More resistant to component faults. Redundant components and double supply lines. Not always resistant to maintenance (probably available during power failures, not available during maintenance).
- Level 3. Most resistant to component faults. All components redundant, double supply lines and backup power. Every component can be maintained without unavailability. Hardware can be distributed over multiple locations. (These components are available during failure).

The reliability of an application is the lowest reliability figure of any of its hardware components. The reliability of a process is the lowest of that of its supporting applications.

The plan then continues with a description of existing countermeasures. Based on the reliability of each process, the analysts can advise further coun-

termeasures and response procedures in case of availability incidents. The reliability figures of processes, applications, software and components help to prioritise countermeasures.

The model continuity plan is useful, as it makes clear which applications and hardware components are important, based on a pre-existing list of critical processes. The plan is therefore closely tied to organisational priorities. However, the analysis part of the framework is rather weak. Analysts do not need to justify their selection of applications or hardware components. If crucial hardware components are overlooked, this will have a significant effect on the overall risk evaluation. Because the reliability of an application is defined as the minimum reliability of its hardware components, omission of even a single low-reliability hardware component will inflate the reliability of the application. Analysts also need not justify their assessment of reliability levels. As an example, the system behind the emergency number 1–1–2 was given the highest possible reliability score (level 3) despite the fact that numerous technical and organisational flaws in this system had been reported [75].

3.2.3 Information technology

Risk assessment methods for information technology tend to focus more on security than on safety or reliability. This is no surprise, as general risk assessment methods for technical infrastructures can be applied to information technology as well but tend to pay less attention to security. Since malicious and remote attacks are especially relevant to information technology, methods for this discipline had to ensure that security risks were well taken into account.

CORAS

CORAS has been developed as a risk-management framework for security critical systems, especially IT systems [45]. CORAS is based on the ISO 31000 framework described in the previous Chapter. Currently, CORAS is presented as a general asset-driven risk management framework [111].

In broad outline, risk assessment in CORAS takes the following steps. First, stakeholders and assets are identified. Assets can be tangible (e.g. equipment items) or intangible (e.g. an organisation's reputation). Harm to one asset can cause harm to other assets; e.g. loss of confidentiality of information may cause loss of reputation. When this has been documented, vulnerabilities and threats to these assets are collected from a gathering of experts from diverse backgrounds. Threats in CORAS include natural, accidental and deliberate causes of potential unwanted events. In order to identify all threats, checklists may be used. When the project leader is satisfied that all threats have been identified and described in sufficient detail, risks are estimated. To estimate the risk posed by threats, the likelihood and consequences of each threat are assessed. Either a qualitative or a quantitative scale may be used. Risk evaluation is

Figure 3.5 – Annotated example of a CORAS threat diagram (verbatim from [111]).

classical, by multiplying likelihood and consequences when assessments are quantitative, or by a risk matrix when assessments are qualitative. For risks that are unacceptable, or that exceed an agreed threshold, suitable and affordable risk treatment options can then be suggested.

A CORAS risk assessment requires input from experts from different disciplines and backgrounds, guided by an experienced lead analyst. CORAS makes extensive use of graphical models, based on the Unified Modeling Language (UML) [78]. Figure 3.5 shows an annotated threat diagram, describing security threats to a server; note that this example uses a mix of quantitative and qualitative assessments. These diagrams describe logical dependencies and cause-consequence relations at a fairly high level. Diagrams in CORAS do not model physical relationships. The CORAS framework is general, and can be applied outside the domain of information technology or security.

ISO 27000 series for risk management

The International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) have jointly published a series of international standards on information security management: the ISO/IEC 27000 series [80]. This series describes information security risk management, a management system, and auditing and certification guidelines.

An information security management system is defined in ISO/IEC 27001 as “that part of the overall management system, based on a business risk approach, to establish, implement, operate, monitor, review, maintain and improve information security. The management system includes organizational

structure, policies, planning activities, responsibilities, practices, procedures, processes and resources.” A management system should be monitored and regularly reviewed, to adapt to changes in environment and discovery of new vulnerabilities.

The 27000 series is generic, describing at high level the required actions and what considerations should be taken into account, without prescribing a particular method. This makes the standards widely applicable, but also leaves it to implementing organisations to choose specific methods. For example, 27001 specifies that “secure areas shall be protected by appropriate entry controls to ensure that only authorized personnel are allowed access.” Standard 27002 gives six implementation guidelines, e.g. stating that visitors should be pre-approved, that a log book should be kept, and that all personnel should visibly wear badges [81]. The 27000 series also contains standards addressing specific sectors, among them standard 27011 tailored to telecommunication service providers [83]. The 27011 standard extends the above implementation guidelines, for example by stating that front desk receptionist should “check the visitor’s belongings at the point of entry and departure in order to prevent him/her from bringing dangerous objects into the premises or taking out assets”.

Several auditing and certification guidelines ensure a minimum level of quality and equivalence between implementations.

The 27000 series has two properties that diminish its applicability to the domain of telecom service availability risks. First, the 27000 series focuses on the process of managing products and services, not on the products and services themselves. It concerns the management of risks in systems and its certification, not the design and building of low-risk systems. An information security risk management system requires a suitable risk assessment method, but the 27000 series does not provide one. The 27005 standard [82] gives guidelines for risk management, and examples of risk identification, analysis and evaluation, but only in non-normative annexes. The series does specify a risk assessment method that can readily be applied by an organisation.

Secondly, the 27000 focuses on security of information, with some attention to its availability, but not so much on the availability of technical components.

3.3 Discussion

The methods from the first and second ages of safety are based on classical risk management. Risk severity in these methods is determined by expected damage, where expectation is used in the statistical sense of probability times effect, summed over all scenarios [50]. The classical approach is limited, in that the expected damage can only be computed when all effects can be expressed quantitatively. Moreover, uncertainties must be low, and neither probability nor effect must have extreme values [97].

Similar restrictions apply to risk assessment methods that were developed for information technology. Methods for this domain tend to focus on logical architecture and deliberate human actions (security), at the neglect of spontaneous failures and physical aspects. Logical architectures describe the function and behaviour of components. These descriptions are suitable for analysing malicious incidents, because attackers have a goal in mind and use functional relationships to select from all vulnerabilities those that help obtain their goal with a minimum of effort and chance of detection. Natural and non-deliberate incidents do not occur with any intent or purpose. Analysing them requires a detailed architectural model describing the physical components and their layout. Risk assessment for security risks therefore requires different models than for safety risks.

Methods from the third age of safety augment classical risk management with socio-technical considerations. These methods are suitable to a wide range of risk problems, and are quite complex. A simpler method tailored to telecommunication systems and crisis organisations may be more efficient.

Current methods in the discipline of crisis management typically focus on high-level risks, and do not analyse telecommunication and information system risks in detail. Some crisis organisations view service level agreements with telecom providers as a suitable mitigation measure for availability risks. This is unsatisfactory, as service level agreements typically exclude the exceptional circumstances in which crisis organisations must operate.

Experience from practice at Radiocommunications Agency Netherlands suggests that many crisis organisations do not perform any structured risk assessment on unavailability of telecommunication services that they use. One possible explanation is that current methods are inadequate. The topic of adequacy is explored in the next chapter.

Requirements

This chapter is based on the author's publications "A new method to assess telecom service availability risks" in the proceedings of the 8th International Conference on Information Systems for Crisis Response and Management, ISCRAM 2011; and "How to assess telecom service availability risks for crisis organisations?" in Advances in Safety, Reliability and Risk Management, the proceedings of the European Safety and Reliability conference (ESREL) 2012; and "Design and initial validation of the Raster method for telecom service availability risk assessment" in the proceedings of the 9th International Conference on Information Systems for Crisis Response and Management ISCRAM 2012.

Research question RQ1 concerns the adequacy of current risk assessment methods. The previous chapter provided an overview of current methods. This chapter has three goals: to define "adequacy" in a way that is useful to our domain, to elicit requirements for risk assessment methods in our domain, and to compare existing method to the stated requirements.

The first section of this chapter addresses the definition of adequacy. This is not a simple task because, as described in Chapter 2, there are different and seemingly incompatible views on the meaning of risk. Risk can be seen to exist objectively, or as something that is socially constructed. In either case, uncertainty is an essential aspect. Probability theory was described as one of many ways to express uncertainty. However, the concept of probability also has different interpretations: frequentist and subjective. Each of these different interpretations affects the meaning of "adequate", and therefore the requirements, but none can be dismissed out of hand. Each appears to have merit in certain cases.

As a path towards solving this definition issue, the first section describes a risk management framework that combines the different interpretations of risk. The key to this combined framework is to distinguish levels of scope in risk assessment. This framework is applicable to any risk management that faces public scrutiny, and is not specific to unavailability of telecom services used by

crisis organisations.

To illustrate how the framework can be applied in the crisis management and telecommunication domain, the second section describes three short case studies.

In the third section, we elicit the requirements for risk assessment methods specifically for this research domain (Section 4.3). As we stressed in the introduction in Chapter 1, practical usability is leading here. This precludes certain approaches that we consider not viable, given practical constraints.

In the fourth and last section, current methods are compared to the listed requirements (Section 4.4). The chapter ends with a conclusion.

4.1 A risk management framework

This section discusses risk management under public scrutiny in general. It introduces risk factors and a framework that shows how risk factors can be used in different levels of scope of risk assessment in order to separate and combine different meanings of risk. The next section will illustrate how this risk management framework can be applied to telecommunication services and crisis organisations specifically.

4.1.1 Arguments for risk estimates

The purpose of risk assessment is to provide stakeholders and decision makers with the necessary information on which they can base and justify their judgement of risk acceptability and choices for risk treatment. A risk assessment must therefore accurately and comprehensively describe the risk.

For the purposes of this chapter, two definitions are introduced. *Formal risk* is defined as that what results when an assessor applies a formal, well described, methodologically sound risk assessment method to the risk target. The application of such a risk assessment method requires skills and knowledge, and only sufficiently qualified assessors are able to compute formal risk. *Informal risk* is defined as what results when an assessor informally estimates or guesses the risk severity. No prior qualifications are required, and for any two assessors, their corresponding informal risk may well be different.

Assessors are defined as the collection of all subjects who perform a formal or informal risk assessment. It is assumed that all stakeholders (subjects who are affected by the outcome of a risk) will make at least an informal risk assessment. Those who perform a formal risk assessment are called *analysts*. The *decision maker* is the person, collective or organisation that chooses risk treatment options. The decision maker does not need to be an analyst, and often is not. The decision maker does not need to be an assessor himself, although this is possible. The decision maker uses the formal risk, in addition to other inputs, such as goals,

availability of resources (e.g. money, manpower), lost opportunities, legal and policy issues, and possibly other factors, including his own informal risk.

The kind of scenarios that are of particular interest to this thesis (unavailability of telecom services) are relatively rare. There is not a large body of identical incidents from which to draw statistical inferences. Furthermore, there are typically substantial knowledge gaps about the causes and mechanisms underlying such scenarios. It is also seldom possible to derive objective probabilities of risks from a structural analysis of known mechanisms. Therefore, the likelihood of risks must almost always be guessed to some extent, based on expert experience and knowledge of causes, mechanisms and effects. Like any other scientific activity, this process is fallible. Probabilities of risks should therefore typically be interpreted as subjective probabilities, not as relative frequencies (as described in Section 2.4.2, “Meaning of probability”).

Similarly, subjectivity not only applies to the probabilities, but also to the effects of risks.

From the fact that a likelihood or effect must be estimated, it does not follow that it is arbitrary. Arguments can (and should) be given why a subjective estimate is reasonable. In addition to structural arguments, analysts can also point to similar previous instances. In absence of disturbing influences, it can be assumed that the event will likely happen in future with the same frequency and impact as in the past. Thus, probabilities and effects of past events can be used to support or “calibrate” future probabilities and effects [94].

Risk scenarios often feature common themes (motifs, tropes), such as irreversibility of effects, involuntary exposure of stakeholders to the risk, or strong feelings of fear by stakeholders. These themes are called *risk factors*. Risk factors can be used both to assist risk analysis and in providing arguments to support a risk assessment. For example, consider a scenario in which an irreversible effect crops up during assessment. For this scenario a certain overall effect had been estimated, but the analysis is fallible and may contain mistakes. If so, the overall actual effect could possibly become much larger than previously estimated, because there is no way to control (reverse or reduce) irreversible damage afterwards. Irreversibility therefore tends to increase the possibility of the effect being much higher than anticipated. As a result, the uncertainty of the effect of the risk tends to increase as well.

Through risk factors, a risk assessment is connected with the arguments justifying it. Obviously, strong arguments are preferred over weaker ones and ideally the argument would be so strong that for all practical purposes it can be treated as truth. A strong scientific examination is therefore required. Some risk factors can readily be examined scientifically, and others less so. For the purposes of this chapter, two extreme kinds of risk factors are defined. ‘*Strong*’ factors are those which can be assessed by models or simulations, and are supported by statistical analysis of past comparable events and situations.

Figure 4.1 – Two faces of decision making. The decision should reduce risk to the risk target as well as reduce criticism about the decision making itself. Arrows denote influence.

Uncertainties can then be modelled using frequentist probabilities. ‘Social’ factors can only be measured indirectly if at all. They can describe human perceptions or intentions, or unfamiliar situations for which assessment has to rely on comparison with analogous but not entirely similar situations. ‘Social’ factors, being based on relatively weak evidence, have higher uncertainty. Only subjective probabilities can be used to describe uncertainty in such factors. The relative merit of ‘social’ factors cannot be derived from a single case, but requires comparing the justifications provided by each factor over a number of cases.

The distinction between ‘strong’ and ‘social’ factors is not a binary classification, but spans a continuous range. Note that ‘social’ factors may be weak in their justifications, but they can nevertheless be forceful and dominate the risk agenda.

4.1.2 Trade-offs in decision making

There are at least five complications in assessing risk and making risk decisions under public scrutiny: the dual nature of risks faced by decision makers, their trade-off between time and accuracy, different approaches to uncertainty and fallibility, and private ownership.

(1) Each stakeholder formulates a (sometimes implicit) list of preferred treatment options, based on his formal or informal risk, goals, priorities, resources, policies, and preferences. Stakeholders may not always accept a difference between their preferred treatment option and the option chosen by the decision maker. Despite the decision maker’s careful selection and best intentions, stakeholders or society at large may raise objections against the decision. These protests may harm the decision maker’s reputation and status, and may lower trust in the decision maker. Criticism may also force the decision maker to choose a different and suboptimal risk treatment. Decision makers therefore face two risks: the formal risk to the risk target, and a informal risk of public criticism (see Figure 4.1). This duality forces the decision maker into other trade-offs.

(2) One of these trade-offs is between accuracy and timeliness. Analysts wish to compute formal risk as accurately as possible. Given unlimited resources all knowledge gaps can be resolved. However, stakeholders may not wish to be exposed to a risk longer than necessary, and may press decision makers for a

treatment choice based on imperfect knowledge.

(3) Decision makers also face the different approaches to uncertainty by the scientific community and the lay public. When two analysts express conflicting scientific opinions, they still view each other's input as additional information that reduces overall uncertainty and that helps in reaching true conclusions. To the lay public, however, conflicting scientific evidence highlights a presumed underlying lack of knowledge. The lay public demand clarity, and when the scientific process cannot provide a clear-cut answer this may be perceived as ineptitude and decrease credibility [95].

(4) In addition, the public is well aware of the fallibility of analysts. Hansson [58] writes: "Insufficient attention has been paid in risk management to the fallibility of experts. Experts often do not realize that for the non-expert, the possibility of the experts being wrong may very well be a dominant part of the risk [...]". Especially for events with low likelihood — such as wide-scale, long-duration telecom failures — the likelihood of errors in analysis exceed the likelihood of the event itself [124]. Decision makers need to rely on experts' recommendations, but must remain critical at the same time.

(5) The last trade-off is specific to the telecommunications industry. Telecom infrastructure is largely owned by private companies. Unavailability risks, however, are very much a public issue. Telecommunication has become vital to so many societal functions that unavailability is no longer an exclusively private concern. When large-scale incidents occur, venting of criticism and discussion about responsibility for the incident are carried out not inside a corporate boardroom, but publicly, in the media or even in parliament.

4.1.3 List of risk factors

Many factors have been suggested that may affect formal and informal risk assessment, and public reaction to risk treatment decisions. The risk factors in *Figure 4.2* have been found in a broad literature review [29, 43, 60, 97, 137, 142, 143, 163]. The factors have been grouped according to their nature.

The first group in *Figure 4.2* contains the factors that concern the physical aspects of the risk scenarios and risk target. These factors are mostly 'strong', and risk treatment can relatively easily influence their values. Classical risk analysis limits itself to this group, specifically to the factors probability, extent of damage and incertitude. The second group contains those factors that reflect directly on risk treatment options. The third group comprises factors that depend on judgements by social groups. These factors are 'social', and do not readily lend themselves to risk treatment. The last group contains all factors that are internal to individuals. These personal attitudes may sometimes be based on a false or imprecise understanding of the risk scenarios and risk target. More than other risk factors, influencing personal attitudes requires providing information and other educational activities. Catastrophic potential

<i>Factor</i>	<i>Description</i>	<i>Effect</i>
PHYSICAL PROPERTIES		
Delayed effects	Characterises a long time of latency between the initial event and the actual impact of damage; the time of latency could be of physical, chemical, or biological nature.	inc
Extent of damage	Adverse effects in natural units such as deaths, injuries, production losses etc.	inc
Incertitude	Overall indicator for different uncertainty components.	inc
Persistence	Defines the temporal extension of potential damages.	inc
Probability of occurrence	Estimate for the relative frequency of a discrete or continuous loss function.	inc
Reversibility	Describes the possibility to restore the situation to the state before the damage occurred (e.g. reforestation and cleaning of water).	dec
Ubiquity, extent of exposure	Defines the geographic dispersion of potential damages.	inc
TREATMENT ASPECTS		
Institutional control	Close, effective monitoring of risks by authorities.	dec
Personal control	Level of control that an individual stakeholder can exercise (e.g. being a driver of a vehicle versus being a passenger).	dec
Voluntariness	Amount of free choice an individual has in being exposed to the risk (e.g. a daily commute versus holiday travels).	dec
SHARED ATTITUDES		
Artificial, immoral	“Unnaturalness” of risk sources (e.g. natural foodstuffs versus artificial additives). Artificiality may also have moral aspects (“tampering with nature”).	inc
Benefits	Tangible and intangible beneficial effects.	dec
Blame	Responsibility for damages clearly attributable to some subject.	inc
Potential of mobilisation	Violation of individual, social, or cultural interests and values generating social conflicts and psychological reactions by individuals or groups who feel inflicted by the risk consequences.	inc

Figure 4.2 – (continued on next page)

Violation of equity	Discrepancy between those who enjoy the benefits and those who bear the risks.	inc
PERSONAL ATTITUDES		
Catastrophic potential	Fear of sudden, disruptive, large effects, as compared to risks that have a small chronic effect over a period of time.	inc
Children involved	Amount of risk exposure faced by children in general.	inc
Dread	Characterises the amount of fear.	inc
Familiarity, known history	Characterises the extent to which the risk is perceived as common and well known.	dec
Media exposure	Amount of attention by the press and broadcasters.	inc
Personal attributes	Gender, level of education, age, income, state of mind, and personal skill level.	–
Trust	Levels of trust in those responsible for the risk.	dec

Figure 4.2 – Risk factors, and their descriptions. In the Effect column, *inc* means that the factor tends to increase (aggravate) risk and *dec* indicates a decreasing (lessening) effect. For personal attributes, females tend to rate risks higher; those with lower education, income or skill level tend to rate risks higher; more anxious people tend to rate risks higher.

is included as a personal attitude, because it is the fear of a catastrophe that is most important here. Its aspects of enormous and irreversible damage are already covered under physical properties.

Several of these risk factors are closely related; for example, Fischhoff et al. [43] show that familiarity correlates with voluntariness, and show other examples of correlation. These factors may therefore be combined, and the number of factors reduced. On the other hand, Vlek and Stallen [163] propose more factors than are shown here, by listing separate components of risk factors. For example, they break the risk factor ‘voluntariness’ down into four constituent parts. Our list of factors could therefore be made shorter or longer. This is akin to “zooming in”, that increases the amount of detail that is visible, but that does not add new information.

Note that factors from the physical properties group are not necessarily ‘strong’ factors in all scenarios. For example, some risks may be so new or ill-understood that probability and extent of damage cannot yet be assessed by measurements or models. However, if in any scenario a factor is ‘strong’, it must be a physical property. Likewise, factors belonging to treatment aspects or shared or personal attitudes will always be ‘social’ to some extent. The scale from ‘strong’ to ‘social’ factors therefore corresponds with the vertical layout of

Figure 4.3 – A risk management framework that combines the different viewpoints on risk and probability.

the table, but only roughly so. With the possibility of these exceptions in mind, the term *social risk factors* will be used to mean ‘all factors other than physical’ in the remainder of this thesis.

4.1.4 Combining different meanings of risk

In this section the concepts presented so far are used to describe a combined risk management framework.

The use and inclusion of risk factors in risk assessment depend on the scope of the assessment (see Figure 4.3). When the scope is limited to just the risk target, only its physical properties need to be heeded. If the scope includes the decision maker and interactions with the risk target, both physical and treatment aspects must be included. In the widest scope, account is taken of all stakeholders, and the shared and personal attitudes of these stakeholders must be taken into account as well.

As discussed in Section 2.3.1, objectivism holds that non-quantifiable (social) factors should not form part of risk assessment. Such an approach shifts the burden of evaluating social factors to decision makers, but provides them with little guidance on how to assess, use, and judge social factors. As one of the problems in our domain is that discussion on treatment decisions is carried out publicly, decision makers need to be provided with strong justifications. Objectivism should therefore be limited to situations where the scope of the risk assessment is limited to the risk target. For this scope, the viewpoint of constructivism is far removed from actual practice as experienced by domain experts. Constructivism is however a helpful viewpoint when the scope of the risk assessment is the interaction between stakeholders, decision makers, and the risk target.

The framework in Figure 4.1 and Figure 4.3 therefore combines the different viewpoints of risk. Each viewpoint provides analytical benefits in certain levels of scope, but would be inappropriate in others.

This chapter does not discuss how decision makers use the information in a risk assessment to reach a decision; it focuses on risk assessment and not on risk treatment. However, several well-known methods exist to assist in qualitative multi-criteria decision making, including AHP (Analytic Hierarchy

Process, [140]) and Macbeth (Measuring Attractiveness by a Categorical Based Evaluation Technique, [22]), and methods based on multi-attribute utility theory or multi-attribute value theory. The applicability of these techniques in our domain is not evaluated in this thesis.

4.1.5 Defining adequacy

Finally, the meaning of “adequate” can be defined. A risk assessment method is *adequate* if it provides decision makers with assessments together with arguments that can be used to justify risk treatment options. The assessments and their arguments are scope-dependent. When the scope is limited to the risk target itself, assessments are based on physical factors: factors that we called ‘strong’. In this scope, risk is objective (objectivism), and probabilities can be relative frequencies. When the scope is expanded to also include the decision maker, treatment aspects become relevant. In the widest scope, when all stakeholders are included, shared and personal attitudes become relevant as well. In the latter two scopes, risk factors are subjective, which we called ‘social’. In these scopes, risk cannot be objective, but is a social construct (constructivism). Probabilities must then reflect a personal degree of belief.

4.2 Cases

In order to illustrate this framework and to show how social risk factors affect assessors and decision makers and the complications that have been identified, three risk scenarios are examined. These are not full case studies, but only serve as case in point of the relevance of risk factors. The cases involve experts as well as the general public, and provide a wide range of causes for unavailability: societal, technical, natural and political. In each case, the role of risk factors is determined in the two risks faced by decision makers. Specifically, cases are used to discover when social factors are most relevant, which risk factors are most relevant, and how these factors can compensate each other.

4.2.1 Case Health risks of electromagnetic fields

In many countries a small group of citizens claims to be sensitive to electromagnetic fields (EMF). They experience headaches, sleeplessness, fatigue and other ailments, and attribute these to EMF exposure. These health effects are of high interest, because EMF are generated for radio-communications, such as cordless in-home telephones, wireless computer networks in homes and offices, and mobile telephony networks. Mobile telephony in particular has drawn attention. Deployment of a mobile telephony service requires the installation of many base stations, each serving a part of the coverage area. Installation of base station antennas has often been opposed by residents in that area, because of health

concerns of EMF exposure. Mobile telephony, and wireless communications in general, are of high interest to crisis organisations.

The risk target in this case is the deployment of EMF in populated areas. Decision makers are the governmental regulators who set exposure limits and license the use of EMF. Decision makers base their decisions in part on risk assessments by scientific advisory bodies. Many studies have been conducted on the biological and health effects of EMF exposure. Literature reviews of these have been published by, for example, [2] and [61]; also see [169]. The body of scientific evidence collected so far supports the conclusion that below the exposure limit there is no causal relationship between EMF exposure and the health issues experienced by some people. A relationship has been found, however, between health issues and the assumption of being exposed. No credible physical mechanism has been found to explain these health effects; the health issues themselves often have no medical explanation.

There are stakeholders who criticise the risk treatment decisions. The BioInitiative Report, for example, is well known for its criticism on current exposure limits [18]. However, this report has been widely discredited by the scientific community [61, 154]. Still, decisions makers are under pressure to reduce exposure limits.

The risk assessments include several risk factors; see for example studies by [27]. Lack of *trust* in scientific consensus and a difference in interpretation of incertitude feature in the alternative views on EMF risks. This lack of trust extends to governments and scientific advisory councils, such as the International Commission on Non-Ionizing Radiation Protection (ICNIRP). *Uncertainty* about the future effects of long-term exposure is greater than for short-term effects. *Delayed effects* can therefore not be ruled out. The alternative view characterises the health effects as pervasive (*ubiquity*), and therefore unavoidable (*voluntariness*) except by extraordinary measures, such as relocating to remote countrysides or shielding houses. Opportunities for *personal control* only arise when tenants have an option to lodge a formal objection to installation of antennas on the place of residence, or refrain from using modern communications. Assessors who adhere to the alternative view often fear EMF (*dread*), using the more suggestive term ‘radiation’ instead of ‘radio waves’ and indicating the possibility of increased cancer risk. Health effects are especially reported for *children* and, to a lesser extent, the elderly and people with ill health. The general public has little knowledge about the physical properties of EMF (no *familiarity*), and the technical design of mobile telephony [28]. The *artificiality* and relative newness of applications such as WiFi and mobile telephony are often remarked upon.

Several social risk factors seem to be highly relevant in this case, including artificiality, involvement of children, dread, familiarity, personal control, trust, and voluntariness.

4.2.2 Case Triple play

Advances in technology allow the same operator to offer three traditional services (television, internet access, and telephony) in a single subscription package. This combination offers greater convenience and cost savings to consumers, and increases customer loyalty for operators. The downside of these so-called triple play subscription packages, is that three important services depend on a single infrastructure, raising the possibility of all three failing simultaneously. When subscribers have no access to television, telephony and internet, they are cut off from the three most important sources of information. When this happens during crisis situations, it will be difficult to warn citizens, provide information and recommend courses of action.

The risk target in this case is the subscription of an individual consumer, who is also the decision maker. The consumer typically does not determine formal risk; most often risk is determined very informally.

Most consumers do not realise their vulnerability, and simply choose the *benefits* of an easy and cheap subscription offer [119]. Very few subscribers fully realise the effect that a large scale outage will have on their lives; *familiarity* with risks of triple-play subscriptions is still low. But equally important is the high degree of *incertitude*. Subscribers do not know the likelihood of failures; operators do not know the effects of failures; there is little experience with outages of triple play services. It is debatable whether customers knowingly, and hence *voluntarily*, take the risk of unavailability. They do, however, have an easy and simple way of avoiding the risks once known, namely to cancel their subscription and revert to previous services (*reversibility*).

Relevant social risk factors in this case include benefits, familiarity, and voluntariness.

4.2.3 Case C2000

TETRA (Terrestrial Trunked Radio) is a technical standard for two-way radio communication between handheld terminals and a central dispatch or other communication networks [153]. It offers secure, simple, and versatile communication within closed user groups. As a result, TETRA has been adopted by many crisis and emergency organisations within Europe and elsewhere. In the Netherlands, the national TETRA implementation for use by crisis organisations is called C2000 [114]. C2000 has come under repeated criticism from some users, because of alleged shortcomings: lack of range, sudden loss of connectivity, poor indoor coverage and inability to communicate, all of which can be dangerous in emergency situations, e.g. when entering a burning building [130]. There have been several incidents in which congestion occurred, e.g. the “Polder Crash” of a Turkish Airlines airplane at Amsterdam Schiphol airport in 2009 [129]. Apart from these congestion incidents, decision makers claim that C2000 performs as designed [117].

In this case, the risk target is the C2000 telecommunication system. The risk in this case is the possible failure of communication during emergency situations. Decision makers are those officials responsible for choosing and mandating C2000; stakeholders are the operational users of the system and, to a lesser extent, the press and the general public.

Risk and risk perception in this case are influenced by many risk factors. The two most important effects or damages resulting from the risk are injuries of crisis personnel and higher damages due to the crisis situation because of less effective crisis response (*extent of damage*). The *probability* of this happening is high, as communications are central to crisis response. There is no *voluntariness* and *personal control* is very limited, since the use of C2000 is mandatory. Several fire services indicated their preference for the previous analogue radios. Risk distribution is felt to be *inequitable*, especially by firemen, who commonly risk their lives in order to help others. Lack of *familiarity* with the principles of TETRA (e.g. that, unlike analogue systems, passive use consumes scarce network capacity) increases the risk of communication failure. The distance between decision makers in government ministries and operational users makes that decision makers a little too easily *blame* users ("lack of ether discipline") and users too easily blame decision makers.

Relevant social risk factors thus include blame, familiarity, voluntariness, personal control, and violation of equity.

4.2.4 Discussion

The cases involved various telecommunication systems and differing causes for unavailability: regulatory restrictions in the case of EMF, technical accidents in the case of triple play, and different expectations between system designers and actual usage in the case of C2000. Nevertheless, in all cases social risk factors contributed to the description of risk; using classical approaches only, the nature of the risk could not have been made clear to decision makers. It can therefore be concluded that classical approaches to risk assessment are not sufficient for telecommunication service availability risks for crisis organisations. The framework presented in the first part of this chapter provides an approach that is better suited to our research domain.

In each case social risk factors were encountered from all four groups. That is to be expected, as in all three cases the scope of the risk assessment was quite broad; it included the technical aspects of the telecommunications system (the risk target), the decision maker, and external stakeholders (citizens in the case of EMF, consumers in the case of triple play, and emergency workers in the case of C2000). Not all risk factors from Figure 4.2 appeared in our example cases (*media exposure*, for example, did not feature). However, the reviewed literature shows that realistic example cases can be found for all risk factors with the exception of purely personal attributes (gender, age, etc.). These attributes are

indicative of risk perceptions at an individual level, but appear to be of little use for risk assessment in telecom service availability. Personal attributes can therefore be removed without loss of applicability in this domain, but other than that there seems to be no need to reduce the list further.

Risk factors can compensate each other's influence on risk. In the EMF case, the one possible factor with a decreasing effect on risk (the benefits of wireless communication) is insufficient to compensate for the many negatively contributing factors. In the case of triple play, possible misgivings about multiple services failing simultaneously have been compensated here by strong voluntariness, reversibility, and (perhaps most importantly) benefits. The C2000 case did not contain attenuating risk factors. Risk factors thus help explain why some risks are considered bad whereas others receive a lukewarm reception. However, it is difficult to predict the net effect of aggravating and mitigating factors. Any factor can in principle be compensated, depending on the particular scenario.

4.3 Requirements of risk assessment methodology

The first section of this chapter provided a framework that combines the objectivist and constructivist viewpoints on risk. In our research domain the scope of risk assessment includes all stakeholders. The framework then shows that in order to be adequate, risk assessments should take all risk factors from Figure 4.2) into account (with the exception of *Personal attributes*, as mentioned). This is one requirement on risk assessments. Other requirements derive from the typical needs of crisis organisations from properties of the current telecommunications industry; these are discussed in the next subsections.

4.3.1 Challenges from crisis management

We already identified one additional requirement: that of practical usability by domain experts. As explained in Section 1.2, crisis management calls for an interdisciplinary approach. Risk assessment in this domain therefore requires the input and participation of experts from various backgrounds. In general these experts will not be experienced risk assessors; however, they will be experts in their particular domain.

A second challenge was mentioned in the Introduction. Modern crisis organisations have built their operational activities around a central, common information store. This increases their dependency on telecommunications. In Section 1.2.4 this was called network centric operations.

4.3.2 Challenges from telecommunications

Ten to twenty years ago telecommunication infrastructures used a vertical model. For example, the telephony service used its own dedicated protocols,

systems, infrastructure and resources. Radio broadcasting also used dedicated systems and resources, etc. Each service formed its own vertical “stovepipe”. The internet’s IP layer, however, has become the dominant carrier for all kinds of services. This leads to telecommunication markets that are more horizontally organised. Providers specialise on basic infrastructure (e.g. dark fiber), IP transport, or services, without necessarily handling the full vertical chain. As a result, services are far less bound to specific transport networks. For example, VoIP telephony can be carried over the copper network, cable networks, GSM/UMTS mobile networks, or private WiFi networks. Web services are also independent, if not completely agnostic, of the transport network used to access them.

Modern telecom services are therefore composed of networks and services operated by many independent and often competing companies. For instance, a phone call may originate at a virtual mobile network operator, using the GSM network of a wholesale operator, that has outsourced its physical network. The information that analysts require as input to their risk assessment has to be obtained from each of these parties. Two complications then arise.

First, a detailed physical model cannot be constructed, for it is impossible in practice to discover the arrangement of all cables, wireless links, routers, switches and other components participating in delivering a service. Each infrastructure itself is a complex network of equipment and connections, often spanning multiple countries. Each network consists of thousands of interconnected items.

Secondly, the companies are not necessarily aware that they participate in delivery of a compound service. For example, a cable may carry data for some essential service without the company that manages the cable knowing about that service. Even if companies are aware of the services, they have little incentive to co-ordinate its availability with competitors. Detailed information on their network is kept confidential, as it would give valuable insights to competitors. It is possible that they understate the availability risks for competitive advantage. The accuracy and reliability of the information that analysts obtain from companies will therefore be uncertain or suspect. Analysts will have to estimate, guess, or conjecture when reliable information is unavailable.

Further requirements emerge from the risk assessment practice in the organisation of the author, Radiocommunications Agency Netherlands. In its operations, the Agency is confronted with society’s increased dependency on telecommunications services. This is evidenced, for example, by the increasing scarcity of available space in the parts of the radio spectrum that are suitable for telecommunications [131, 132]. As a result, two issues become more important. First, there is great interest in incidents that cause outage, and telecom risks in general. Second, and because of this interest, explanations on policy decisions and justification of the Agency’s actions are addressed no longer exclusively to experts, but increasingly to non-experts, such as politicians and concerned

citizens. Furthermore, private and public sector organisations alike have a pressing need to reduce costs; there are few resources for in-depth studies, unless they yield results that can be applied immediately. For these various reasons, the Agency can no longer limit itself to isolated elements of telecom systems, such as the radio path only, but has to consider the delivery of telecom services from end user to end user.

4.3.3 Initial requirements

Based on these considerations, three groups of requirements were identified: to handle the challenges (R1–R3), to ensure ease of execution (R4–R6), and to guarantee the usefulness of the results of the method (R7–R9). These requirements are called ‘initial requirements’, to indicate that they form the starting point of the design of the new method. Experiments, case studies and field tests may lead to improvements of the method, and indeed have. These improvements are described in detail in the next chapters; below the initial requirements are listed.

R1 *No full model*: the method must not assume the entire telecom system to be modelled in detail.

This requirement derives mainly from the first challenge: complexity prevents a full model to be created in practice. But even without practical constraints, telecom companies would still be unwilling to disclose confidential information, for reasons of competition. Modelling of the risk target should therefore allow for knowledge gaps on the architecture of the telecom networks.

R2 *Uncertainty*: the method must take into account uncertainty about the correctness of input information.

Information on services and network architectures is not only incomplete, but may also be unreliable. Companies may overstate the reliability of their services, and may understate risk. Furthermore, because large telecom outages are relatively rare and because telecom infrastructures are continuously updated and modified, there exists no body of relevant and representative past incidents. Statistical inferences on likelihoods and effects are therefore difficult, if not in many cases impossible.

R3 *Social risk factors*: the method must be able to take into account a diverse range of social risk factors.

The risk management framework presented in the first part of this chapter showed that as the scope of the assessment expands additional risk factors should be taken into account. Some of these risk factors are subjective, or need to be estimated based on weak data. Nevertheless, these risk factors are informative to decision makers, and may influence risk mitigation choices and priorities.

- R4 *Usability*: Experts from different fields (telecommunications, crisis management, engineering) must be able to understand and apply the method.

Telecom experts are able to understand the failure mode of technical components and mitigating factors, but only crisis management experts can indicate the impact that this failure will have. Other experts bring their legal knowledge or knowledge of organisational structures to the table. A team of experts from mixed backgrounds must be able to perform the risk assessment. This implies that for practical applications the risk assessment method must support and facilitate collaboration.

- R5 *Effort*: it must be possible to execute the method in an acceptable amount of time.

Experts' time is scarce and expensive, and a method that is time consuming is therefore unlikely to be used in practice. Limits of what is considered acceptable in practice will be investigated during the final field tests (Chapters 8 and 9).

- R6 *Exceptional circumstances*: the method must be able to take account of large-scale disruption and other exceptional circumstances that likely arise during crisis situations.

It is not sufficient to base the risk assessment on quiet situations between crises. Among other things, this implies that service level agreements are likely not a sufficient basis by themselves, as service level agreements typically exclude disasters and other exceptional circumstances.

- R7 *Identify all risks*: the new method must identify all causes of unavailability that are within the scope of the risk assessment.

Decision makers require some certainty that all relevant risks have been taken into consideration. Full certainty means that all unexpected risks must be included, which is impossible by definition. Nevertheless, decision makers need an argument why there are no risks that should have been taken into account, other than the ones that have been assessed.

- R8 *Over-aggregation of information*: although risk assessment may be based in part on subjective evaluations, it should always be possible to retrace the reasoning that led to a particular assessment of risk.

The new method must therefore not summarise or hide key information, and the effects of any uncertainties in the input information must be reflected in the risk assessment. Arguments supporting the assessments must be available, so that the decision maker can verify their soundness.

- R9 *Priorities and treatment options*: the new method must support the determination of risk priorities and the selection of preferred risk treatments, suitable to this research domain.

As stated previously, the purpose of any risk assessment is to provide stakeholders and decision makers with the necessary information on which they can base and justify their judgement of risk acceptability and choices for risk treatment. Figure 4.1 showed that decision makers in this research domain face two risks: that to the risk target, and the risk of public criticism. The risk assessment method should support the choice for risk treatments that are both effective and supported by the public.

4.4 Current methods versus the requirements

This section compares existing methods against the requirements. The same order as used in Chapter 3 is used here as well: methods from the disciplines of engineering, crisis management, and information technology respectively.

Engineering. The methods from the first age of safety limit the scope of their risk assessment to the technical components of the target. This is insufficient for our needs: requirement R3 states that social risk factors should be taken into account as well. This rules out second-age methods as well, but leaves third-age methods as possible options. However, third-age methods such as STAMP and FRAM are complex (R4 *Usability*) and time consuming (R5 *Effort*). In summary, methods from the engineering discipline are not immediately applicable, and difficult to adapt to our requirements.

One approach would be to extend one of the engineering methods, to adapt it to our requirements. Since detailed information on the telecom infrastructures is lacking (R1 *No full model*), first-age methods such as FTA cannot be applied. FTA has other disadvantages, such as difficulty in handling software and common cause failures. The analytical parts of the FMEA method can be of use but, again because of requirement R1, the system (the end-to-end telecom service) cannot easily be decomposed into layers and elements as required by FMEA. Highly quantitative methods cannot be used because accurate data is not available (R2 *Uncertainty*).

Crisis management. The discipline of crisis management and risk governance yields useful insights, but few practical methods that can be applied to our domain. The continuity plan framework is in active use, but has severe analytical limitations; it does not satisfy R7 *Identify all risks* nor R8 *Over-aggregation of information*. It is a method that will be useful in combination with a thorough risk assessment. Methods from the crisis management discipline provide us with examples of how a method could treat ambiguity, uncertainty, and social risk factors, but not with risk assessment methods that can be applied or adapted to our requirements.

Information technology. Methods from the discipline of information systems typically model the information dependencies between components, and understate the relevance of physical components and hazards in an end-to-end

communication chain (*R7 Identify all risks*).

None of the current risk assessment methods fully satisfies the requirements. This sufficiently justifies the development of an alternative risk assessment method, one that is tailored to our problem domain. The method proposed in this thesis is not entirely new. It contains elements from FMEA and other methods, with features to handle incomplete and missing information, and social risk factors.

4.5 Conclusion

Decision makers in our domain need risk assessments that are at the same time scientifically sound and deserving of public approval, as they need to balance both the risk to the risk target and potential public criticism. Strong risk factors are required for scientific rigour, and social factors are required to co-opt public approval. A scientifically well-grounded risk assessment may indicate effective mitigation actions, but if they would cause a public outrage then these will not be implemented, or cause new problems. On the other hand, a risk assessment merely based on public opinion but without scientific justification, may lead to ineffective mitigation actions supported by public opinion. Effective risk assessment must consist of both a scientific assessment of the risk target as well as an assessment of public response. The descriptions of risk factors can be provided in three parts: those that reflect to physical properties of the risk target and the scenarios, those that reflect to treatment options, and those that reflect to stakeholders who may be affected by the risk outcome. By presenting these three descriptions to the decision makers, analysts equip them for risk treatment that is both effective and supported by the public.

Furthermore, a risk assessment methodology for our domain must take the specific properties of this domain into account. Models of telecommunication services will be both complex and incomplete. Busy experts from very different educational and professional backgrounds need to collaborate. Risks to cover range from relatively common failure scenarios to singularly rare and catastrophic failures. Finally, the method must effectively assist decision makers to choose risk treatment options, and to explain those choices to all stakeholders. A full list of initial requirements was enumerated.

Currently available methods from the disciplines of crisis management, engineering and information technology do not satisfy these requirements. This justified the development of a new method.

A design science approach will be used to create a new methodology that meets these requirements. In the next chapter, a first version of a methodology is described. In later chapters this first attempt is improved upon, based on the results from validation tests.

Initial Design of the Raster Method

Parts of this chapter have previously been published in the author's publication "A new method to assess telecom service availability risks" in the proceedings of the 8th International Conference on Information Systems for Crisis Response and Management, ISCRAM2011. This chapter is based on the initial draft of the Raster application manual.

Chapter 1 announced that this research would follow a design science methodology. In particular, the engineering cycle would be used to repeatedly improve our artifact, which is in this case a risk assessment method. In the previous chapters a problem investigation was performed and design requirements were specified. This chapter describes the first actual design, to meet these requirements. In subsequent chapters this design is changed and improved after validation tests, and several features described in this chapter have disappeared or been modified in the version of the RASTER method described in Appendix D. Notes in the inside margin mark changed or discontinued features.

5.1 Telecom service models

To allow experts from various domains to work with the method, and to allow them to collaborate despite differences in knowledge, a risk assessment method is proposed that is based on graphical models. RASTER models telecommunication services as a graph of typed components. Five node types are used: actors, wired links, wireless links, equipment, and unknown links. The example in Figure 5.1 models satellite telephony as used by a small crisis team. To make the diagrams more meaningful to non-technical participants, nodes can be decorated with icons representing the physical component. The node types are described first; their connection rules are described thereafter.

Actors represent users of the telecommunications service. Actor nodes may

The Raster tool does not support decoration

Figure 5.1 – Early example of a telecom service diagram: satellite telephony as used by a small crisis team. Slanted rectangles indicate equipment items; wireless links are indicated by rounded rectangles; cloud shapes indicate unknown links (see text).

be a single individual or a group of individuals with a common role. Sometimes a distinction is made between the main actor and co-actors. The main actor is a member of the crisis organisation, and uses the telecommunications service in this professional capacity. Unavailability of the telecommunications service to the main actor is therefore the focus of our method. The co-actors may be members of the same or some other crisis organisation, or may not be affiliated with any crisis organisation. They may, for example, represent the press or the general public.

Wired links represent passive physical cables, including connectors and joints, but without intermediate active devices. These can, for example, be fiber-optic cables, analogue telephone copper pairs, or marine cables. Cables always lead from one node to one other node; these nodes are not part of the wired link, and are modelled separately. A cable containing multiple strands or fibers is often best modelled using a single wired link. For separate cables in the same trench or duct, it may be more convenient to use multiple links.

Wireless links represent radio-frequency connections, without intermediate devices, such as public broadcasting networks, point-to-point links, or shared channels used by handheld two-way radios. The transmitter and receiver(s) themselves (and their antennas) are not part of the wireless link, and are modelled separately. A wireless link can connect more than two nodes when they can share the same radio frequency. Alternatively, point-to-multipoint links can be modelled as multiple wireless links, one to each multipoint.

Equipment nodes represent any known physical component of a telecommunications service that is neither actor nor a link, such as switches, exchanges, or computing equipment. Equipment does not need to correspond to a single item or a single equipment cabinet; it could correspond to an entire installation. For example, it is possible to model a small telephone exchange as a single

equipment component.

Unknown links represent subsystems of the telecommunications service that are not modelled for sake of conciseness, or that currently cannot be modelled because there is too little or no knowledge. Unlike wired and wireless links, which represent single physical channels, unknown links represent a collection of wired and wireless links and equipment.

In early publications about the RASTER method, the terms ‘unexplored link’ and ‘unresolved link’ were sometimes used instead of unknown link. We finally settled for unknown link, because that term can be used both when we do not wish and when we cannot expand the link in more detail. ‘Unexplored’ and ‘unresolved’ might be taken to mean that exploration is the intention or rule, whereas one sometimes purposely refrains from expansion. Calling a node an unknown link only indicates that the subsystem is not modelled in detail, regardless of whether that is by choice or by necessity.

Both equipment items and unknown links are abstractions for more detailed collections of physical entities. Unlike unknown links, for equipment items it is known that no further subdivision is possible, or that no subdivision is required in order to complete the risk assessment. Using unknown links a model can be built for any telecommunication service. This model will be complete in scope, but may not yet have enough detail for risk assessment.

A telecommunications service model is well-formed if it conforms to certain rules. These rules follow from physical limitations. Nodes correspond to physical entities, and these rules correspond to their physical restrictions. A well-formed graph thus represents a possible physical reality.

1. The model forms a simple, connected graph (there will be no edges from a node back to itself, no multiple edges between two nodes, and no disconnected partitions).
2. The model must contain precisely one main actor and at least one co-actor. Each actor must be connected to at least one piece of equipment or unknown link (actors cannot connect directly to a wired or wireless link).
3. Each wired link connects to exactly two pieces of equipment or unknown links.
4. Each wireless link connects to at least two pieces of equipment or unknown links.
5. Each equipment node must be connected to at least two other nodes of any type excluding other equipment (equipment components must not be connected to each other).

Later relaxed to ‘at least one’.

In the remainder of this thesis, all models are assumed to be well-formed.

The most simple model consists of two actors connected by an unknown link. (Technically, two actors connected by a single equipment item is also

Figure 5.2 – Model refinement. Node u is replaced by a set of new nodes and edges.

Figure 5.3 – Shapes used in diagrams to indicate the type of nodes.

well-formed, but this can hardly be said to model a *telecommunications* service.)

Any telecommunication service model that contains an unknown link can be refined. By refinement extra detail is added to the model; the lack of knowledge that is represented by the unknown link is reduced. When a model is refined, a node is replaced by a set of new nodes and edges between those nodes that jointly have the same interface (see Figure 5.2). This refinement of models is called *expansion*. The same physical reality is described by both the old model and its refinement. However, the new model — having more nodes and edges — can capture this physical reality in more detail.

Requirement R1 *no full model* implies that expansion of unknown links should not always be necessary. Any method using telecom service models and diagrams must therefore stipulate clear rules to decide when expansion of an unknown link is unnecessary.

Models of telecommunication services are graphs, and can therefore be expressed graphically. In our notation, the shape of a node indicates its type (see Figure 5.3). Actors are represented by a face symbol, indicating that they represent human end-users of the telecommunication service. Equipment items are indicated by a slanted rectangle. Since equipment items are hardware, the symbol use sharp (‘hard’) corners. Wireless links are indicated by a rectangle with rounded edges, suggestive of the ‘soft’ and intangible nature of radio waves. Wired links are indicated by a plain rectangle. Like equipment items, cables are hardware and have sharp corners. Unknown links are indicated by a cloud symbol.

In this thesis the term *node* is used as well as *component*. ‘Node’ means a graph element of a telecom service diagram. ‘Component’ means the physical entity represented by the node. Nodes do not have vulnerabilities; components do. The same component can be represented by more than one node, if that component is represented in more than one diagram.

Figure 5.4 – Conceptual model of risk analysis using RASTER (initial version) as a UML class diagram. Attributes in regular font are assessed by the analysts; attributes in italics are computed.

5.2 Risk analysis in telecom models

The models described in the previous section correspond to a single telecommunication service. An organisation typically uses several services (e.g. fixed and mobile telephony, internet connections, etc). A single RASTER project therefore consists of multiple diagrams. Each diagram describes a single service, and consists of multiple nodes that implement the service. This conceptual model is shown in Figure 5.4. The figure shows the initial version; the model will be improved in the next chapter. The list of most urgent risks for the entire project is ultimately based on the assessments of vulnerabilities on nodes and node pairs (for common cause failures), as explained in the next subsections.

5.2.1 Use of risk factors

Chapter 4 investigated risk factors in assessment of telecom service availability risks. Personal attributes were already excluded in the discussion on risk factors. Based on discussions with telecom experts, we suspect that of the physical risk factors, Delayed effects are negligible for evaluation of vulnerabilities in telecom service diagrams. All other risk factors listed in Figure 4.2 are included in our evaluation scheme. This scheme uses two values, called Frequency and Impact, to assess the physical risk factors. The *Frequency* value combines the risk factors incertitude and probability of occurrence. The *Impact* value combines the risk factors extent of damage, incertitude, persistency, reversibility, and ubiquity. The use of non-physical risk factors is explained in Section 5.2.3.

Our motivation for having two values is to strike a balance: sufficiently simple and easy to use, yet with values that are verifiably accurate. Other methods sometimes split the likelihood of an incident into likelihood of the threat and quality of the defences and protective measures. We chose not to do so, since this would require detailed information on the infrastructure, which is commonly not available (see R2 *Uncertainty*); the extra complexity outweighs the doubtful gains in accuracy of assessments. The odds of some disaster striking the nodes in the telecom infrastructure, or the effectiveness of protections, are

Final definitions: see Figure 6.5.	Class		Frequency
	H	High	Once every 5 years
	M	Moderate	Once every 50 years
	L	Low	Once every 500 years
	U	Extremely low	Immeasurably low
	V	Extremely high	Immeasurably high
	A	Ambiguous	When experts disagree
	X	Uncertain	When uncertainties are too high
	–	Undetermined	Not yet assessed

Figure 5.5 – Frequency classes for telecommunications service vulnerabilities. The Frequency score includes the risk factors *probability* and *incertitude*.

seldom known. Instead, past incidents can be observed and a Frequency value derived from them.

Both Frequency and Impact are divided into the same eight value-classes. The scoring system in Figure 5.5 is used for Frequency; Impact is scored using a similar scoring system in Figure 5.6. Each class is abbreviated by a single letter or symbol: U for extremely low (Unlikely or Ultra low), L for Low, M for Moderate, H for High, V for extremely (Very) high, X for Uncertain, A for Ambiguous, and – for undecided.

Three classes (High, Moderate, and Low) are used to describe cases when Frequency and Impact can be assigned a more or less accurate position on a numerical scale. Two classes (Extremely Low and Extremely High) are used to describe cases for which the classifications Low, Moderate and High are inappropriate because of extreme values. Together, these five classes are called *definite* classes. The classes do not describe a range of values, but give a single typical, representative value. For each value, the analysts decide on the most appropriate class. Definite classes have low to modest uncertainty, such that it does not prevent the assignment of a single class to the factor. For example, a value of Medium means that despite any uncertainty on the exact value of the factor, it is clear that its value falls within the class Medium as opposed to High or Low.

Three additional classes describe likelihood and impact with a high Incertitude. If there is lack of consensus among the analysts about the most appropriate class, then that factor is marked as *ambiguous*. If there is lack of knowledge but no explicit disagreement, then the factor is marked as *uncertain*. The default class is *undetermined*, meaning that no attempt for classification has yet been completed. This class is present so that it is possible to distinguish between assessments that could not be completed because of lack of knowledge, versus assessments that simply have not been looked at yet. These three classes are called *indefinite* classes.

The typical values given for the classes low, moderate, and high are working

Class		Impact	Final definitions: see Figure 7.7.
H	High	Complete, prolonged unavailability of the service	
M	Moderate	Partial, temporary unavailability of the service	
L	Low	Noticeable degradation of the service	
U	Extremely low	Immeasurably low	
V	Extremely high	Immeasurably high	
A	Ambiguous	When experts disagree	
X	Uncertain	When uncertainties are too high	
–	Undetermined	Not yet assessed	

Figure 5.6 – Impact classes for telecommunications service vulnerabilities. The Impact score includes the risk factors *extent of damage*, *incertitude*, *persistency*, *reversibility* and *ubiquity*.

assumptions, which need to be verified in further studies.

When estimating the likelihood of a vulnerability, RASTER takes the likelihood of disasters into account (see R6 *Exceptional circumstances*). For example, if the disaster “flooding” occurs once every 75 years, and flooding leads to an incident once every two cases, then the likelihood of that vulnerability is once every 150 years. The analysts will then assign either class “moderate” or “low”.

Note that the impact is not limited to the node to which the vulnerability belongs. Impact is assessed by its overall effect on the telecom service; vulnerabilities are local, but their impact is service-wide. Still, only effects that influence the telecommunications service are taken into account; consequential damage (damage to other parties) is not included in the impact classification.

5.2.2 Evaluation of vulnerability score

From the Frequency f and Impact i of some risk the vulnerability score of that risk can be computed. The operator $\otimes$ denotes the vulnerability score, as $f \otimes i$. For the outcome of these scores the same set of classes and symbols is used as for Impact and Frequency, and with the analogous meanings.

Our calculation table for $\otimes$ is given in Figure 5.7. It is based on the following considerations. We want the part of the table for Frequency and Impact L, M, or H to match expected damage. These three classes represent modest values, for which classical risk assessment is considered to be suitable. RASTER thus encloses classical risk assessment as a special case. Expected values are computed by multiplying probability and effect, i.e. the values for Frequency and Impact. In the RASTER method, Frequency and Impact are not numerical. However, it is easy to find a mapping from numerical value ranges to values classes L, M, H and vice versa, such that the product of the numerical equivalent of Frequency f and Impact i maps to the numerical equivalent of $f \times i$.

For the other two definite classes, U and V, our considerations are as follows. When impact is extremely high, it does not matter what the Frequency is, as the risk is unacceptable at any likelihood. When likelihood is extremely high

Definition
changed:
see Fig-
ure 6.7.

Impact	A	A	A	A	A	A	-	A	
	-	-	-	-	-	-	-	-	
	X	X	X	X	X	X	-	A	
	V	U	V	V	V	V	X	A	
	H	U	M	H	H	V	X	A	
	M	U	L	M	H	V	X	A	
	L	U	L	L	M	V	X	A	
	U	U	U	U	U	U	X	A	
	⊗	U	L	M	H	V	X	-	A
		Frequency							

Meaning of letter codes: U stands for extremely low (Unlikely or Ultra low), L for Low, M for Moderate, H for High, V for extremely (Very) high, X for unknown, A for Ambiguous, and - for undecided.

Figure 5.7 – Combining Frequency and Impact into a single vulnerability score (initial design).

(i.e. near certainty), it is almost certain that damage will arise, and we are therefore obliged to prepare countermeasures. In this case the risk will also be unacceptable. When the impact is extremely low (i.e. nearly absent), we do not really care whether the incident happens; the risk will always be extremely low to us. The same consideration applies for situations where the likelihood is extremely low. Therefore, $f \otimes i = U$ when either f or i equals U, and $f \otimes i = V$ when either f or i equals V (and neither equals U).

Changed:
see Sec-
tion 6.5.4.

Looking at the part of the table describing the combination of definite values, it can be noted that when either the likelihood or the impact is not known, the combination also cannot be known. Therefore, when either f or i has an indefinite value, the result should also be indefinite. In this combination, we always want to preserve ambiguity, as we believe that information to be highly relevant to decision makers. When an undetermined value is involved, the result must also be undetermined as that value could turn out to be ranked as ambiguous rather than simply unknown; until the value of that factor is assessed, the result of the $\otimes$ operation is still undetermined. When neither the value A nor - is appropriate, the combination is ranked as a 'plain' unknown (symbol X). This concludes our consideration of the calculation table.

The left and right operand of $a \otimes b$ have different physical meanings: Frequency and Impact. It is by coincidence that $\otimes$ happens to be commutative ($a \otimes b = b \otimes a$). In particular, commutativity means that $L \otimes H = H \otimes L$. In practice a low-likelihood/high-impact vulnerability is very different from a high-likelihood/low-impact vulnerability. Stakeholders typically prefer the right variant (near certainty of low impact), as any insurance salesperson knows. At first glance, RASTER does not seem to make a distinction. Remember, however, that the classes L, M, H are for moderate values only, where the expected damage will be an adequate measure of the vulnerability score. When either Frequency

Figure 5.8 – Adjustment of the overall vulnerability level based on social risk factors. Arrows indicate the eight permitted adjustment steps.

Later dropped; became part of Evaluation stage.

or Impact is extremely high, their combination will yield V, which carries a meaning much different from L, M, H. Furthermore, even when Frequency and Impact are non-extreme, the combination into a vulnerability score is only a first step in risk assessment. When the vulnerability turns out dominate the overall risk assessment, a more careful consideration of all its hazard scenarios will become necessary. At that time, the initial assessment of likelihood and impact will still be available for review.

Also note that according to our definition $\otimes$ is idempotent ($a \otimes a = a$). This is also by coincidence, not by requirement, and alternative definitions may be possible that are not idempotent (and still fulfil the requirements).

5.2.3 Evaluation of overall vulnerability level

From the vulnerability scores of some node, the overall vulnerability level (sometimes called the overall risk level) of that node can be computed. This is done by taking the maximum of the values. To do so, a total ordering is placed on the scores, where

$$U < L < M < H < X < - < A < V$$

Changed: see Section 6.5.4.

The motivation for this choice is our assumption that at most one hazard scenario leads to an incident at any one time. The risk level of a node then equals the worst vulnerability for that node.

In the calculation of the overall vulnerability level, the non-physical risk factors are included. These factors are described and noted during assessment, when relevant. There is no fixed assessment scheme for these risk factors as there is for likelihood and impact. Based on the descriptions, the analysts can decide to adjust the overall vulnerability level. If the net effect of the risk factors is deemed to decrease the risk, the next lower level is chosen; if their net effect increases the risk, the next higher level is chosen (see Figure 5.8). The analysts decide by consensus whether risk adjustment is necessary. Risk class A (ambiguous risks) cannot be modified by risk factors. Ambiguity is not a matter of opinion, and can only be resolved by further research. Risk class X (uncertain risks) also cannot be modified by risk factors. Since the position of the risk in the risk-space is unknown, any change in location would still result in an uncertain risk.

Later moved to Evaluation stage.

Figure 5.9 – Overview of the initial method, showing the stages and their inputs and outputs. Arrows denote input and output; the dotted arrows indicate stepwise refinement.

5.2.4 Telecom service risk evaluation

RASTER does not use an aggregate risk level per diagram (telecom service) nor for an entire project. Based on the overall vulnerability assessments of all nodes the analysts determine which telecom services and which parts of their architecture are the most risky. The analysts then review those vulnerabilities, to ensure that the best possible assessment is made, and to propose the most effective treatment options. From the assessed vulnerabilities, the analysts choose by consensus which ones have the highest priority. For these most important risks, they propose treatment recommendations. The final output of RASTER is not an assessment of the organisation's risk level, but a list of proposed risk treatments.

In general, diagrams do not explicitly show the communications path between actors. Connections between nodes indicate physical connections, not the flow of information between participants in a conversation. Telecom service models describe a physical view of telecom services, not a logical one. It is therefore not possible to derive the risk of service interruption from the overall vulnerability levels of nodes.

5.3 Execution of the Raster method

Risk assessment in the RASTER method consists of four consecutive stages: preparation, refinement, common cause failures, and evaluation. See Figure 5.9.

Later 'Refinement' renamed to 'Single Failures'.

In the *preparation stage*, the crisis organisation and its context are described to facilitate the modelling and analysis. This description establishes the goal and boundaries of the risk assessment. It includes a list of disaster scenarios and checklists with frequently appearing vulnerabilities. The list of disaster

<i>Node type</i>	<i>Vulnerabilities</i>
Equipment	Physical violence, such as fire, flood, or earth quake. Power failure; loss of electricity supply. Failure, because of ageing or unpredictable failures. Electromagnetic interference (unexpected EMC issues)
Wired link	Breaks, e.g. due to trenching or falling trees. Congestion, exceeding of transmission capacity. Electromagnetic interference.
Wireless link	Blocking and shielding of radio waves. Congestion, exceeding of transmission capacity. Electromagnetic interference.

Figure 5.10 – Basic checklists of vulnerabilities. These checklists are defined during the Preparation stage.

scenarios limits the effort involved in applying the method. For example, if the disaster scenarios exclude dike breaches in a particular area, the risk of flooding of equipment cabinets in that area can be ignored for the purpose of this risk assessment. The checklists are used to aid risk identification. A basic set of vulnerabilities for the checklists is shown in Figure 5.10. In the preparation stage the analysts add vulnerabilities to the checklists, or remove some, based on their expectation of which vulnerabilities will be most common during risk analysis. Checklists are not limiting; during assessment of individual components additional vulnerabilities can still be considered.

During the *refinement stage*, a telecommunication service model is created and analysed for each telecommunication service used by the crisis organisation. For each telecommunication service, the initial model is drawn according to existing knowledge of the service. It will likely include one or more unknown links. The remainder of this stage is then iterative. In each step one unexamined non-actor node is analysed; risks associated with actors are not taken into account.

For nodes that represent equipment, wired links or wireless links, the analysis describes likelihood, consequences and social risk factors. For each vulnerability of the node, the factors Frequency and Impact are assessed, according to the definitions in the previous section. The vulnerability score of each vulnerability is calculated. Based on assessment of social risk factors, the overall vulnerability level can be adjusted, as described in a previous section. From these vulnerability scores the maximum is chosen as the overall vulnerability level of the component. The checklists should be consulted to ensure that all relevant hazards have been considered.

For unknown links no checklists are used. Instead, all types of vulnerabilities — those that apply to equipment, wired links, and wireless links — are

Later social factors moved into Evaluation stage

analysed as the subsystem may contain such components. If no reliable analysis can be made due to lack of knowledge (when the overall vulnerability level is unknown), the link must be expanded, and hence described in more detail.

In the *common cause failures stage*, sources of common cause failures (CCFs) are identified and analysed. CCFs are multiple failures that share a single underlying cause. We distinguish identity and proximity CCFs. Identity CCFs arise when the same node is present in the models of two different services. Failure of that node could be a common cause for the failure of those telecommunication services. The first sub-step in this stage is therefore to identify all nodes that exist in multiple models. Proximity CCFs arise when two nodes are physically sufficiently close together for them to be affected by the same incident. For example, two equipment items within the same room could be destroyed together in a fire; two cables within the same duct could be damaged by a single trenching incident; two wireless links could be affected by the same jammer. The second sub-step in this stage is therefore to identify all combinations where such vulnerabilities may arise. As in the previous stage, further refinement of unknown links may be required. Common cause failures are described and classified in the same way as single vulnerabilities, as described above.

Node
classes in
the Raster
tool re-
place iden-
tity CCFs.

Later done
through
creation
of clusters.

Later de-
scription of
this stage
much ex-
panded.

In the *evaluation stage*, all results from the previous stages are collated. The analysts determine the overall magnitude of each identified risk and rank them. For risks with a magnitude above some threshold, treatment options and their associated cost are described as well. This ends the risk assessment.

5.4 Discussion

This chapter specified the initial treatment design: the RASTER method for assessment of telecommunication service availability risks. In this section some of the limitations of our design are discussed, and the design is compared against the requirements that were specified in Chapter 4.

5.4.1 Limitations of telecom service models

Telecom service diagrams (or “diagrams” for short) can completely model a telecom service, meaning that all physical components that make delivery of telecom services possible can be represented in the diagram. Diagrams can show any level of detail, depending on the need and on the amount of knowledge that is available. If the knowledge is complete, the diagram will not contain unknown links.

Our diagrams also have limitations. The first limitation is that diagrams are static; they do not model changes over time. The telecommunications industry is highly dynamic, driven by rapid innovation and market changes. Telecom service diagrams act on a snapshot of the current state of affairs. This limits

their usefulness in situations where the intervals between changes are short compared to the time it takes to perform a risk assessment.

Another limitation is that diagrams disregard the dependence of telecom services on performance monitoring and maintenance. Regular monitoring and maintenance is necessary for the undisturbed operation of telecom components. Monitoring may detect deviations before they lead to service unavailability; (preventive) maintenance is necessary to uphold the mean time between failures of components. A complicating factor for crisis organisations is that disturbances with critical infrastructures during crises may hinder (corrective) maintenance work. Floods, blockage of transportation systems, and discontinuity in the supply of food or drinking water will make it difficult for maintenance staff to carry out their duties. Maintenance staff also need telecom services themselves; unavailability of one telecom service may cascade onto other telecom services.

A third limitation is that RASTER's diagrams do not show control relationships among nodes. Examples of such relations are time synchronisation requirements between equipment nodes, or fail-over relationships between redundant components.

Finally, and related to the above, diagrams do not explicitly show the flow of information from actor to actor. When the graph contains multiple paths between actors, additional information is necessary to determine which path or paths actually relay messages.

5.4.2 Limitations of the Raster method

RASTER does not include vulnerabilities of actors. This does not imply that issues with human and organisational reliability cannot be handled by RASTER at all. For example, operator mistakes that cause an exchange to misbehave, or failure to recharge mobile equipment can both be modelled as vulnerabilities of equipment. But other issues cannot be handled by RASTER, such as misunderstanding or lack of cooperation between actors. Organisational resilience and the safety culture of the organisation, among others, are outside the scope of the method.

The present method deals with single failures, and with multiple failures sharing a single common cause. Simultaneous independent failures are not considered, even though their effect may be very serious. The reason for this is the assumption that failures are rare, and the likelihood of non-common cause failures is therefore so low that it can be disregarded.

Because telecom service models describe a physical view and not a logical view, it is not possible to compute the risk of service interruption from the overall vulnerability level of nodes. RASTER also does not compute an aggregate risk level per telecom service or per project. The design of RASTER prioritises practical, actionable results leading to a reduction of risk, over measuring risk. As a consequence it is more difficult to compare the risk levels of organisations, or to measure the change in risk level of an organisation over time.

5.5 Design validation

This section provides arguments to support the statement that RASTER satisfies the nine requirements listed in the previous chapter. In a way, this is a preliminary treatment validation to the validations that are described in the next chapters.

- R1 *No full model*: The use of unknown links and their refinement allows for the creation a model that is complete, but does not contain more information than necessary for the risk assessment.
- R2 *Uncertainty*: RASTER explicitly takes account of reliability in information and disagreement among experts (ambiguity). Limited amounts of uncertainty on the assessment of likelihood and impact are accounted for by the use of value classes, instead of numeric value estimates. This is a natural result of using a coarse grained qualitative scale.
- R3 *Social risk factors*: Social risk factors are explicitly accounted for in RASTER in the analysis of vulnerabilities. All social risk factors can be included; the current method does not cover a specific list and mechanism.
- R4 *Usability*: Although this is a requirement that can only be fully validated in experiments, the use of diagrams intends to improve usability. Diagrams are typically more easily understood than, for example, formal or mathematical models.
- R5 *Effort*: This, too, is a requirement that can only be fully validated in experiments. The use of unknown links can significantly reduce the size and complexity of diagrams, reducing the effort required. Expansion is postponed as long as possible, hoping that in the end it will turn out to be unnecessary at all, thus saving time and effort.
- R6 *Exceptional circumstances*: One of the input documents to RASTER is a list of disaster scenarios. These scenarios are explicitly taken into account during the analysis stages.
- R7 *Identify all risks*: Analysts use the checklist and their experience to enumerate failure causes for each component. They follow a structured methodology, but discovery of all possible failure causes is not guaranteed by the method.
- R8 *Over-aggregation of information*: The results after the evaluation stage can be presented in a condensed form for easy readability, but all underlying assessments are still available. If necessary, every treatment recommendation can be traced back to likelihood and impact assessments of individual vulnerabilities of individual nodes.
- R9 *Priorities and treatment options*: A risk ranking is created in the evaluation phase. It is assumed that risk treatment options for these classes are in most cases obvious, as they will often be limited to simply accepting the risk or stop using the service. Risks with high uncertainty in their risk factors

require discussion among all stakeholders. Possible treatment options include (temporarily) halting the use of that telecom service, or further studies.

Making Raster Work

This chapter is based on the author's publication "Design and initial validation of the Raster method for telecom service availability risk assessment" in the proceedings of the 9th International Conference on Information Systems for Crisis Response and Management ISCRAM2012.

Our initial design needs to be validated. It needs to be demonstrated that our risk assessment method yields the desired results when applied to crisis organisations. The results need to be acceptable, satisfy the requirements, and be reliable. This requires a number of validation experiments. This chapter describes the first such experiment: a case study at the crisis organisation of Radiocommunications Agency Netherlands, to verify whether the method can be practically applied. Based on the lessons learned from this case study, the RASTER method is improved; the modifications and improvements to the method are described as well.

6.1 Research method

In broad outline, this case study validated the initial version and redesigned it based on initial case experience. Validation was continued with the improved risk assessment method. This process was repeated two more times, continuing the case study every time with an improved version of the method. This methodology of incrementally designing and validating a technique is described in more detail by Wieringa [164]. To evaluate the case study, the participants were asked to complete a questionnaire. Also, final interviews were held with individual participants.

6.1.1 Research questions

The only research question at this stage is whether the method can be practically applied. Validations of its usefulness and completeness of the results will be done

in later case studies, as these require a workable method first. ‘Applicability’ is operationalised by three questions:

1. Are the analysts able to apply and complete the method based on the manuals provided to them?
2. Does the method yield justifiable treatment recommendations?
3. Are the analysts satisfied with the results and the time needed to complete the method?

6.1.2 Case description

The case study was performed at the author’s institution: Radiocommunications Agency Netherlands. This is a Dutch government agency responsible for operational activities and enforcement of regulations pertaining to telecommunications. The agency is located in two offices, separated by about 180 kilometres. The internal crisis organisation of the agency forms part of the broader crisis organisation of the Ministry of Economic Affairs. Outside crisis response phases, the Agency’s crisis team advises on usage and technical aspects of telecom infrastructures. During crisis response, the Agency’s crisis team monitors the radio spectrum, locates and resolves interference sources that affect wireless communications, and advises decision makers on technical issues concerning telecom infrastructures.

Five analysts participated in the study. Two of the analysts were active members of the crisis management team; all had experience with technological aspects of telecommunications systems. There was no pre-agreed project plan, but the responsible managers supported the case study, and allowed the analysts to spend some of their time on the study.

At the start of the case study, the analysts found that nine telecom services were being used in crisis response: mobile telephony, fixed telephony to the public telephony network, internal ‘voice over IP’ (voip) telephony between the two office locations, the national emergency telephony system, email, satellite phones, pagers, office automation via remote desktops, and a video conferencing link between the two offices. To reduce time demands on the analysts, they decided to select two services for further analysis: internal voip telephony, and satellite phones. The analysts chose these two, because they deemed them to be the most complex and the least known.

6.2 Case study execution

The analysts held about a dozen two-hour meetings for the case study. The author was present at most meetings as a facilitator, but was not involved in actual execution of the method.

Iteration 1

In the first meeting the facilitator introduced the analysts to the RASTER method, and to the object of the case study. The analysts received a written description of the method for reference. From the first introduction onwards, the analysts offered alternatives to the method and possible improvements to it. Their attitude was critical, but supportive. The analysts then collected information for Stage 1 (preparation). They found that collecting the required information was easy, as most of the information had already been described for their own internal use in a format suitable for the case study. The analysts divided themselves into two teams, one team in each office location. Each team resolved one of the two preselected telecom services.

During Stage 2 (refinement), it quickly became apparent that tool support was needed. The analysts tried to create and record the diagrams and the scoring of vulnerabilities using common office automation software, using templates provided as part of the documentation. This proved to be cumbersome, as information on diagrams and vulnerabilities of components was kept in disconnected applications. When adding or removing diagram nodes, the list of vulnerabilities had to be updated manually.

The case study was therefore paused, and a browser-based tool was created using standard web technologies and common Javascript libraries (see Appendix C). The analysts volunteered in testing the tool, and offered many suggestions for improvement.

Iteration 2

After the RASTER tool was introduced, the case study restarted at Stage 2 (refinement). During Stage 3 (common cause failures), the RASTER tool automatically presented the list of potential common cause failures (CCFs) that had to be assessed. This list was compiled by generating all triplets of a vulnerability and two components where both components are subject to that vulnerability. From this list, only a small fraction turned out to be realistically vulnerable to a CCF. Out of 351 component pairs, the analysts considered only 14 (4%) to be relevant; they concluded that for all other pairs no CCF was reasonably possible. Although the method appeared to yield correct results, the analysts found the procedure too cumbersome and time consuming, thus violating requirements R4 *Usability* and R5 *Effort*.

We then looked at alternative approaches for the CCF stage. At first the tool was augmented with options to quickly dismiss unrealistic pairs. This reduced the workload somewhat, but did not address the core of the issue, as analysts still had to consider all pairs, even if it was just to click the button to disable an unrealistic pair.

Many component pairs can be dismissed out of hand because the geographical distance between them is too large to make their common failure

probable or possible. We considered adding location information to nodes, as well as a maximum effect-distance to each vulnerability, so that the tool could automatically reject component pairs for which a CCF was impossible due to their distance. For example, a fire would only affect equipment within the same building, whereas flooding would affect equipment in a much larger area. By allocating ‘fire’ an effect-distance of 50 metres and ‘flood’ an effect-distance of 30 kilometres, the tool could automatically compute all potential pairs of components for which a CCF is possible. This idea was rejected, because we expected that location information would be difficult to obtain, and that the additional complexity would make the method impractical. Furthermore, not all vulnerabilities are physical in nature. Configuration errors, for example, can affect two components regardless of their physical separation. Proximity in this case means “being maintained by the same support team”. For vulnerabilities such as Ageing, proximity is determined by replacement policies determined by ownership or responsibility.

Instead, an alternative was adopted that was suggested by the analysts themselves. Components are grouped into clusters that represent geographical areas, such that CCFs are confined to components within each cluster. The analysts expected that they would be able to quickly and easily create such clusters, based on their knowledge of the approximate physical placement of components. Instead of a set of geographical coordinates, it would be sufficient to know the region in which the component was to be found. Clusters can be created for other proximity-measures as well. For example, clusters can be created based on the skill level of operators, so that configuration errors can be treated using the same clustering method. The property on which a cluster is based is called its *critical property*. Again, the case study was paused while the tool was updated to reflect this new approach to CCFs.

Iteration 3

The case study was continued and completed in a single full-day session, using the latest version of the method and the tool. Stages 1 and 2 could be reused without any modification, so the case restarted at the revised Stage 3. The analysts created node clusters, and performed their analysis.

The initial design of the RASTER method allows for assessment of social risk factors for each vulnerability of each component. This proved to be difficult as well as impractical. Many of the social risk factors apply to telecommunication services or to the crisis organisation as a whole, less commonly to individual components and hardly ever to a specific vulnerability on a component. The instructions for Stage 4 (evaluation) had been expanded by this time, as it had become clear that a more detailed description of the steps was necessary. The assessment of social risk factors was moved from the refinement and CCF stages into the evaluation stage. The feature of vulnerability level adjustment (see

Figure 5.8) was discarded as a result.

The improved version of the evaluation stage starts with the selection of a longlist. The longlist consists of a selection of vulnerabilities, as assessed in the previous stages. Typically the vulnerabilities with an overall level of Very high and High will be selected, but the analysts are free to choose a different minimum level, and to add or remove risks as deemed necessary. The risks of the longlist always have the form of a vulnerability and a component, e.g. “power failure on the telephone exchange”. This longlist is a draft of the list of risks that will feature in the final risk treatment recommendations of the method. This draft is reduced to a shortlist, in a consensus-based procedure in which the analysts take into account additional information not present in the diagrams. For example, this can include information on control dependencies between components, information about redundancy of components, or the relevance of telecommunication services to the organisation. Services are not of equal importance: high risk of unavailability in a useful but non-essential service may be less important than a moderate risk of unavailability in a top-priority service. This is also the point in the improved method where social risk factors are assessed. Each of the risks on the long list is checked for relevance of each of the social risk factors. The result of these deliberations is not only a reduced shortlist, but also a prioritisation of this list. On the basis of this prioritised shortlist risk treatment recommendations can be formulated. This ends the evaluation stage.

The analysts created the longlist and shortlist for Stage 4 (evaluation). Time constraints prevented them from completing Stage 4 in full, but they did briefly describe the risk factors for treatment aspects and attitudes for each risk on the shortlist. They also discussed possible risk treatments.

6.2.1 Questionnaire and final interview

After completion of the case study, each analyst individually completed a questionnaire (see Appendix B.1). The goal of the questionnaire was to obtain data with which the method could be improved for analysis of single and common cause failures.

The questionnaire contained two main questions. The purpose of the first question was to validate the relevance of each of the physical risk factors. For each vulnerability identified in the case study the analysts were asked about the influence of each physical risk factor. With 20 kinds of vulnerabilities and the 7 physical risk factors from Figure 4.2 they had to score 140 combinations. For each combination, three answers were possible:

- Relevant, meaning that for this vulnerability (e.g. power failure) the assessment of the factor (e.g. persistency) has a considerable or (very) large impact on severity of the risk.
- Not relevant or less relevant, meaning that for this vulnerability the factor

has little or no impact on the severity of the risk, regardless whether the factor is assessed as high or low.

- Blank, when the participant felt unable to answer.

The purpose of the second question was to validate the typical values used in the classes high, moderate, and low (Figures 5.5 and 5.6). At the end of the case study the author selected eight risks (component–vulnerability combinations), four from each telecom service. These eight were selected to give a close to equal number of high, moderate, and low scores for likelihood and impact. The analysts were then asked to estimate upper and lower bounds to the value of each physical risk factor, for the four risks from the telecom service that they analysed. For example, for reversibility they would give an upper and lower bound to the percentage of damage that would be permanent for that risk; for persistency, they would give an upper and lower bound to the time to restore the service for that risk. Note that the Frequency and Impact assessments as previously scored by the participants were not repeated in the questionnaire. Participants had to answer without reference to their previous assessment. It is also unlikely that participants had memorised their previous scores, as Stage 2 (refinement) contained nearly one hundred of these assessments for each team, and three months had elapsed between completion of Stage 2 and the questionnaire.

Finally, semi-structured final interviews were held with the analysts individually. A list of standard questions was used, but free-form answers were allowed and any additional remarks from the analysts were recorded.

6.3 Results and discussion

Because of confidentiality it is not possible to provide detailed results. In Stage 1 (preparation) the checklist with common vulnerabilities contained 7 vulnerabilities to equipment items, 4 to wired links, and 4 to wireless links. In addition to these default vulnerabilities, the analysts added 7 specific vulnerabilities to individual components.

A summary of the results of Stage 2 (refinement) of the two telecom services can be found in *Figure 6.1*. Note that none of the vulnerabilities were assessed as Ambiguous.

In Stage 3 (common-cause failures) the analysts created 41 node clusters, for 17 vulnerability types. Hence, in the common cause failures Stage 41 node clusters were analysed. For stages 2 and 3 together, a total of 217 vulnerability assessments were completed.

During the evaluation stage, a longlist of 26 risks was compiled: 21 single failures and 5 common cause failures. Of the single failures, 13 were ranked as unknown, 5 as high, and 4 as extremely high. Of the common cause failures, 3 were ranked as high, and 2 as moderate.

Figure 6.1 – Summary over the two telecom services in the first case study: number of nodes per type, and number of risks per level (single failures only).

6.3.1 Case study execution

The analysts estimated to have spent 40 hours on the entire case study, and about half that time on executing the RASTER method itself.

The analysts did not rank any vulnerability as ambiguous. They viewed ambiguous scores mostly as a failure of discussion of the underlying reasons for scores. This does not necessarily mean that ambiguity is not a useful concept, but it does suggest that further research may be necessary.

Although 71% of all vulnerability types originated from the checklists, the checklists accounted for 95% of all analysed component–vulnerability combinations. Added vulnerabilities were typically used for a single component only. From this it can tentatively be concluded that checklists are a suitable method for vulnerability identification and help save time but that additional vulnerabilities may be relevant.

It was also observed that the analysts seldom referred back to the descriptions of the seven likelihood and impact classes after having read the manual. This was also reflected in the outcome of the questionnaire, and will be discussed further in the next subsection.

A related observation is that the analysts did not strive for firm justifications of all vulnerability assessments. Their scores were often based on general experience with telecom hardware, not with the properties of specific components. For example, the analysts assessed the likelihood of breaks in Ethernet cables as ‘low’ without searching for specific local historical data to back up this claim. On the one hand this shows a benefit of RASTER, in that it does not require all input data to be equally accurate, but in some cases (e.g. those in the questionnaire; see next subsection) it may have a significant effect. It is likely that with a larger group there would have been more interaction, that estimates would have been more accurate, and that missing data would have been flagged sooner.

Interestingly, the RASTER tool showed at a glance that the internal voip telephony service was rated as less risky than the satellite telephony service.

Factor	Analyst					Overall
	1	2	3	4	5	
Extent of damage	+	+	+	+		+
Probability	+	+	+	–		+
Reversibility	+		+	+		+
Ubiquity	+	–	–	–	+	
Persistency		+	+	–	+	
Incertitude		–	–	+		
Delayed effects		–	–	–		–

Note: + indicates that the factor is counted as meaningful, – that it is counted as not meaningful, and blanks mean undecided.

Figure 6.2 – Questionnaire question 1: relevance of each physical risk factor, by analyst.

Although part of this difference can be explained by differences between the two services (voip telephony may well have higher availability than satellite telephony), the analysts did acknowledge the possibility of an uneven bias between the two self-assigned teams.

6.3.2 Questionnaire and final interview results

The questionnaire contained two main questions.

Question 1. The first question asked, for each combination of 20 vulnerability types and 7 physical risk factors, about the relevance of that factor for assessments of that vulnerability. The answers therefore consisted, for each analyst, of a 20-by-7 table, containing one of three scores: relevant, not relevant, or blank (abstain).

These scores were processed as follows. For each factor the percentage of vulnerabilities that were scored as relevant and as not relevant was computed. Because of abstentions, these percentages do not necessarily add up to 100%. Vulnerabilities were weighted; those that were assessed for many components received a higher weight, and contributed more towards the overall result for that factor. We then count a factor as meaningful if it relevant for at least 15% of vulnerabilities *and* irrelevant for at most 85% of vulnerabilities. A factor is counted as not meaningful if it is relevant for at most 15% of vulnerabilities *and* irrelevant for at least 85% of vulnerabilities. In all other cases we consider the relevance of the factor to be undecided.

The results are shown in *Figure 6.2*. In this table the symbol + indicates that the analyst’s scores were counted as meaningful; the symbol – indicates that the factor was counted as not meaningful; blanks indicate an undecided outcome. The final column shows our summary of the overall outcome. The questionnaires confirmed our assumption that the physical risk factor ‘Delayed effects’ is hardly relevant in telecom service availability risks (unlike for example in health risks of food additives). Questionnaire results unsurprisingly indicated

<i>Service</i>	<i>Vulnerability</i>	<i>Component</i>	<i>Frequency</i>	<i>Impact</i>
satphone	Fault	Up/downlink Burum	L	M
satphone	Congestion	PSTN	L	M
voip	Break	Fiber-optic	L	H
voip	Congestion	Internal network A	M	L
voip	Power failure	Internal network B	M	M
voip	Power failure	Layer 3 switch	M	H
satphone	Physical impact	GPS satellite	H	L
satphone	Power failure	Satphone handset	H	H

Figure 6.3 – Questionnaire, vulnerability assessments selected for question 2. L,M,H mean Low, Moderate, High respectively.

that the physical risk factors probability, extent of damage, and reversibility (repairability) are highly relevant, but did not conclusively show the relevance of ubiquity and persistency. For crisis organisations, the difference between failures with a long repair time and unrepairable failures is not relevant, since crises are typically short-lived. Incertitude was not clearly counted as meaningful. On the one hand this seems surprising, as incertitude is an essential ingredient of risk. On the other hand, the analysts experienced no problems in reaching consensus during the case study, scored very few vulnerabilities as Unknown, and none as Ambiguous.

Question 2. The second question asked for estimates of each of the physical risk factors in specific risk situations, in the form of a range of values (lower and upper limit). This question was asked for four selected risks (combinations of a vulnerability and a diagram component) separately. The selection is shown in Figure 6.3; participants only answered the question for the four risks from the telecom service that they previously worked on. This selection was chosen because it approaches an equal number of High, Moderate, and Low assessments for both kinds of assessments. These previous assessments were not shown to the analysts; they were not reminded of their previous answers, and were asked to give upper and lower bounds to physical risk factors individually rather than for combined risk factors in the form of Frequency and Impact.

The purpose of this question was to validate the typical values used in the classes high, moderate, and low. For Frequency, the typical values at which these classes are anchored are “once per 5, 50 and 500 years” respectively (Figure 5.5). The analysts’ estimates of physical risk factors could confirm the suitability of these anchor values, or provide support for different values and spread of values. Similar reasoning could be applied to the analysts’ estimates for risk factors underlying the Impact classes and the class descriptions in Figure 5.6.

Results of the second question showed an large variation in upper and lower bounds. For example, for the probability of breaks in a fiber-optic cable one analyst answered “once per 10 to once per 50 years”, while an other analyst’s range was “once per 40 to once per 200 years”. On the extent of damage these

analysts disagreed even more: the first answered that “50% to 100%” of the service would become unavailable, while the second answered “5% to 25%”. In general almost all answers showed a similarly large amount of disagreement.

The estimates also showed no consistent relation with the typical values specified in Figures 5.5 and 5.6. There was no correlation between the Frequency-score actually assessed in the case study, and the risk factor probability as estimated in the questionnaire. Similarly, there was little correlation between the Impact-score assessed in the case study and the estimates for extent of damage, ubiquity, persistency or repairability. For comparison, the Frequency and Impact scores in the actual case study were replaced with scores based on the estimates as answered in the questionnaires. In all but one case this led to a higher overall vulnerability level for that component and, using the line of reasoning the analysts employed, to more risks on the longlist, and possibly the shortlist as well.

Testing accuracy and usefulness of the end results were not goals of this case study. However, in future case studies such imprecision would be more serious, as it would lead to inaccurate results when these assessments are compared in the evaluation stage.

Because the results of question 2 did not allow us to improve the scales for Frequency and Impact we decided to discard the results.

We can only speculate on the reasons for this. Perhaps the question was poorly worded, or too difficult for the participants to answer. Perhaps there was too long an interval between the analysts first assessing the risk and them answering the questionnaire. Of course, there is also a possibility that, for whatever reason, reliability of expert assessments in this experiment was very low. Reliability of the method was not a goal for this experiment; the issue is addressed in Chapter 7.

Final interviews. The final interviews made clear that the analysts require more clarification on the method and guidance on its execution. In particular, they indicated that they require more guidance in the assignment of classes to Frequency and Impact. Several points for improvement were discussed; see Section 6.5.

One explicit topic in each interview was the time required for execution of the method. The opinions of analysts ranged from “relatively little” to “quite a bit”. Overall they did not expect to need less time for future studies, as they expect to spend most time on discovering the components and their interconnections. Whether the time required for performing a RASTER analysis is acceptable in practice will be discussed with the results of the field tests, in Chapters 8 and 9.

6.3.3 Research questions

Finally, this section returns to the three questions to determine applicability of RASTER, as listed in the Research method section.

1. Are the analysts able to apply and complete the method based on the manuals provided to them?

With tool support and with the latest version of the RASTER method, the analysts were able to use and complete the method. The analysts did not analyse each of the nine telecom services, but selected two. The effect of this on validity will be small, as they selected the ones that appeared the most complex and least known to them. However, it is possible that use in larger projects will lead to the discovery of specific aspects of the method that make RASTER impractical for projects of that size.

2. Does the method yield justifiable treatment recommendations?

Although time constraints prevented the completion of Stage 4 (evaluation) in detail, the analysts were able to derive a shortlist and discuss suitable treatment commendations. The results have not been presented to an independent decision maker, but the analysts selected those vulnerabilities with the highest overall vulnerability level, identified some aggravating social risk factors, and provided multiple treatment recommendations based on their domain knowledge.

3. Are the analysts satisfied with the results and the time needed to complete the method?

Based on the final interviews, the analysts were mostly satisfied with the results. The analysts found the overall amount of time reasonable, but did not expect to need less time in future. We have no reason to believe that they answered other than truthfully, as they were critical on other aspects such as the lack of useful documentation. Note again that testing the accuracy and usefulness of the end results were not goals of this case study.

6.4 Lessons learned

The case study gave us valuable starting points for three practical guidelines for applying the RASTER method: group size, division of tasks, and method organisation. These lessons learned are used to improve the design (see Section 6.5).

Group size. A group of five analysts participated in this case study. Since the analysts decided to split up into two groups, each group addressing one telecom service, the actual number of analysts involved in analysing each service was less. For purposes of validating the method, having five analysts was sufficient. However, for a full application of the method it would be beneficial if a larger group could be employed, in order to enlarge the available skill set and to promote discussion. The required skill set depends on the particular organisation. In general, it should cover expertise on technical aspects (telecommunications and information technology), expertise on crisis management, and domain knowledge (expertise on the operations of the organisation). Team composition

will be discussed further in the field tests (Chapters 8 and 9). The optimal group size probably depends on the organisation, the distribution of skills, and the complexity of the telecom services. For further case studies, it would be preferable to have at least eight analysts participating.

Division of tasks. Based on the results of this study, it cannot be concluded whether it is beneficial to split the group into sub-groups, each analysing one telecom service. Although this may appear to be more time-efficient for Stage 2 (single failures), there are two reasons to think that it makes other stages less efficient. First, Stage 3 (common cause failures) is done for the project as a whole, not for each individual service; each analyst would therefore need to understand each telecom service in order to participate in the analysis of common cause failures. Secondly, a consistent estimate of likelihood and impact is required in Stage 4 (evaluation). With separate groups analysing each service, it will be more likely that an uneven bias leads to inaccurate evaluation results. However, creation of the initial diagrams could perhaps be done by different subgroups in parallel.

Organisation of the method. The manuals and organisation of RASTER need further thought. More guidance is needed on the scoring system for likelihood and impact of vulnerabilities. A meeting secretary or facilitator can help ensure the quality of the assessments and can help maintain consistency in scoring.

6.5 Design improvements

To complete the case study, we had to apply a large number of design changes, all helping to make the method into one that can successfully be executed from beginning to end within reasonable time. This has resulted in a method that could be performed at least in this practical case. Such was exactly the purpose of the case study. This section summarises the changes that were made to the method, i.e. the changes to the initial design.

6.5.1 Conceptual model

The design improvements lead to a modified conceptual model (*Figure 6.4*) with two changes. The first change is that RASTER now uses node clusters instead of node pairs. This was explained in the description of Iteration 2. Clusters can be nested: a cluster can contain child-clusters as well as diagram nodes.

Secondly, similar components can now share one assessment of vulnerabilities. The RASTER tool supports node classes: there can be multiple instances of a nodes class, and all instances share the same assessment of vulnerabilities. Node classes reduce the amount of work, for example when two identical routers are employed in a redundant configuration; only one assessment needs to be made which is then shared between both routers.

Figure 6.4 – Conceptual model of risk analysis using RASTER (final version) as a UML class diagram. Attributes in regular font are assessed by the analysts; attributes in italics are computed. Compare to Figure 5.4.

6.5.2 Tool support

One major outcome of this case study is the creation of a software tool to assist in creating diagrams, recording assessments, and computing risk levels. In addition to drawing and recording, the tool facilitates analysis, consistency checking, and collaboration between technical and non-technical experts; it also serves as a research tool. The use of common web technologies makes rapid prototyping easy. The tool was extensively modified from its initial concept throughout this case study, and has continued to evolve based on feedback during subsequent experiments with the method. Tool development is not one of the research goals, but was needed to conduct the research. In the research, the tool is a way to collect data about the use of the method; in practice, the tool is needed to use the method at all. Initially, the descriptions of the RASTER method have been kept independent of any tools, including the RASTER tool itself. During the course of further improvements to both the tool and the method it has become more difficult to maintain this strict separation. Executing the RASTER method without any tool support has become practically impossible, as the RASTER tool makes Stages 2 and 3, and parts of Stage 4 a lot faster and easier. The current version of the RASTER manual assumes that the RASTER tool is used. Note especially that validation of the tool has not been a specific research goal, although the tool was used during the two final field tests.

6.5.3 Telecom service diagrams

The initial design specified rules for well-formedness of diagrams. These rules referred to physical limitations of components. One of these rules was that equipment items need at least two connections to components of any type excluding equipment. For example, a handset could be connected to an actor (as its user) and to a cable connecting the handset to the wall. During the case study a use case was discovered for which this rule is too strict. A server (e.g. a database server or mail server) can legitimately be connected to an office

network with a single connection. The rules on diagrams were relaxed to allow for this use case.

The telecom service diagrams initially allowed for the inclusion of small decorations representing the physical appearance of the component. The diagram in Figure 5.1, for example, shows pictures of satellites and a ground station. Although possibly useful, this feature was not present in the versions of the RASTER tool used in this case study, because the design and programming efforts for this functionality seemed overly large, given the merit of the feature.

6.5.4 Risk analysis

Several improvements have been made to the risk analysis prescriptions of the method.

Definition of Frequency and Impact descriptions

From the interviews and from observations during the experiment it was learned that the current descriptions in Figure 5.5 of High, Moderate, and Low frequencies cause confusion. The descriptions were found to be too abstract. For example, if a component is replaced every five years, the Frequency of a vulnerability can still be ‘once every 50 years’. Additional explanations and tips were added to the RASTER manual to clarify the meaning of these phrases. Also, some diagram nodes do not represent a single specific component but stand for a number of components, such as the handsets of all crisis responders in general. This is a useful shortcut, as there is little to be gained by repeating a large number of identical nodes in the diagrams. The interpretation of ‘once every 50 years’ was found to be confusing in these cases as well. To ease interpretation of High, Moderate, and Low frequencies in these cases, an alternative description was added that applies more clearly to multiple cases. These meaning of these alternatives is identical; the incident likelihood is merely restated in different words.

To meet the general request for more guidance, the descriptions for the other classes and for Impact were elaborated as well. The improved descriptions are given in *Figure 6.5* and *Figure 6.6*.

Evaluation of vulnerability score

Experience within the working environment of the author suggest that low-probability, high-impact events are an important source of incidents. Cases in point are the Waalhaven incident in 2011 [74], and the incidents with the emergency number 1-1-2 in 2012 [75]. Potentially harmful events with low likelihood but high impact are difficult to identify. These risks are easily dismissed as “unrealistic” or even “impossible”. Taleb has coined the term ‘black swans’ for these events [151]. Unless black swans are identified, one cannot begin to assess and mitigate them. Once dismissed, the risks will not

<i>Class</i>	<i>Frequency</i>
H High	Once every 5 years, or: for 1000 identical components, each month 15 will experience an incident.
M Moderate	Once every 50 years, or: for 1000 identical components, each month 1 or 2 will experience an incident.
L Low	Once every 500 years, or: for 1000 identical components, each year 2 will experience an incident.
U Extremely low	Very rare, but not physically impossible.
V Extremely high	Routine event. Very often.
A Ambiguous	Indicates lack of consensus between analysts.
X Unknown	Indicates lack of knowledge or data.
– Undetermined	Not yet assessed

Figure 6.5 – Improved Frequency classes.

<i>Class</i>	<i>Impact</i>
H High	Long-term, but eventually repairable unavailability of the service for all actors.
M Moderate	Partial temporary unavailability of the service for some actors.
L Low	Noticeable degradation of the service.
U Extremely low	Unnoticeable effects.
V Extremely high	Very-long term or unreparable unavailability of the service for all actors.
A Ambiguous	Indicates lack of consensus between analysts.
X Unknown	Indicates lack of knowledge or data.
– Undetermined	Not yet assessed

Figure 6.6 – Improved Impact classes.

feature in further risk analysis, and will therefore not influence the risk treatment recommendations.

It is therefore important that the RASTER method handles black swans gracefully. The instructions on removing vulnerabilities from components were changed to explicitly state that vulnerabilities should only be removed if they are physically impossible, not just merely unlikely. A typical example is the vulnerability of ‘flooding’ on a space satellite, or ‘configuration mistake’ on a device that does not allow for any configuration. In the RASTER method, black swans receive the Frequency Extremely low and Impact Extremely high. In the initial version of the method (see Figure 5.7), if either Impact or Frequency is Extremely low the vulnerability level is defined as Extremely low; if either Impact or Frequency is extremely high the vulnerability level is defined as

Impact	A	A	A	A	A	A	A	A	
	-	-	-	-	-	-	-	A	
	X	X	X	X	X	X	-	A	
	V	A	V	V	V	V	X	A	
	H	U	M	H	H	V	X	A	
	M	U	L	M	H	V	X	A	
	L	U	L	L	M	V	X	A	
	U	U	U	U	A	X	-	A	
	⊗	U	L	M	H	V	X	A	
		Frequency							

Meaning of letter codes: U stands for extremely low (Unlikely or Ultra low), L for Low, M for Moderate, H for High, V for extremely (Very) high, X for unknown, A for Ambiguous, and - for undecided.

Figure 6.7 – Combining Frequency and Impact into a single vulnerability score (improved design). Compare to Figure 5.7.

Extremely high. These rules are ambiguous for black swans. However, there is a vulnerability level for ambiguity. The combination rules for Frequency and Impact were updated to this effect.

One other change to this table is the case when Frequency is undecided and Impact is scored as Ambiguous. In this case, the vulnerability level can have no other result than Ambiguous, regardless of how the Frequency will be scored. In Figure 5.7, however, the vulnerability is considered still to be undecided in this case. (Likewise for the case when Impact is undecided). This mistake was corrected in the combination table as well. The new combination table is shown in Figure 6.7.

Evaluation of overall vulnerability level

The rules for evaluation of the overall vulnerability level (see Section 5.2.3) were modified as a result of the changes made to accommodate black swans. Vulnerabilities are now removed less often: only if physically impossible, not when merely unlikely. As a result, more assessments would have to be completed by the analysts. This is because the old ordering

$$U < L < M < H < X < - < A < V$$

implies that the overall vulnerability level of a component will remain undecided (indicated by -) until each and every likelihood and impact of vulnerabilities of that component have been assessed. (Except, of course, if a vulnerability is scored as Ambiguous or Extremely high (A or V), which is rare). Before, if a vulnerability was judged to be less important it would simply be removed from the list of vulnerabilities of that component. Because of changes due to black swans (in the previous subsection) this is no longer allowed. Still, some vulnerabilities could justifiably be regarded as less important. To allow such

Figure 6.8 – Final overview of the method; compared to Figure 5.9 only the names of the stages have changed. Arrows denote input and output; the dotted arrows indicate stepwise refinement.

vulnerabilities to be left unassessed, thus saving the analysts from unnecessary effort, the evaluation ordering of vulnerability scores is changed to

$$- < U < L < M < H < X < A < V$$

With this updated order, it becomes effectively optional to assess vulnerabilities on components. If none of the vulnerabilities on a component are assessed, the overall score will be -; this was true also in the initial design. Whenever at least one vulnerability is assessed, the vulnerabilities that are assessed determine the overall score. This overall score can change when more vulnerabilities are assessed, so the original ordering was in a sense more correct, but less useful in practice.

6.5.5 Execution of the method

During the case study and in the final interviews the analysts indicated that several aspects of the RASTER method needed further clarification. As a result, several sections of the RASTER manual were rewritten and extended, to provide more explanations and guidance. Chapter 7 of the RASTER manual (see Appendix D) is now entirely about practical aspects of performing the RASTER method, and includes a discussion on division of tasks (when this is possible, and when it is not advisable).

Also, Stage 2 was renamed from 'Refinement' to 'Single failure analysis' to better cover the goal of that stage. Refinement can also be done as a result of Stage 3. The new outline of the RASTER method is shown in Figure 6.8.

6.6 Conclusion

From these results some general conclusions can be drawn.

First, if a risk assessment method is based in part on a model that is updated or refined during the assessment, then tool support is a practical requirement in order to maintain consistency between model and assessments, and among assessments. Without proper tool support, any changes to the assessments due to changes in the model have to be recorded by hand. For all except toy examples, this quickly becomes unworkable in practice. In addition, tool support is needed to maintain consistency between assessments. When adding components to the model, additional and previous assessments must be similar if the risks are similar. The same finding applies to, for example, STPA; as more practitioners apply STPA to practical cases, the need arises for tool support for STPA as well [1].

Second, consistency of assessments by domain experts cannot simply be assumed, but has to be induced by the method by explicit design. It is not sufficient to state in the method's manual that experts should ensure that their assessments are consistent. The method itself should provide feedback mechanisms or review steps to detect and correct dissimilar assessments in similar situations, and similar assessments in dissimilar situations.

Testing Raster's Reliability

This chapter was previously published as “Experimental Validation of a Risk Assessment Method” in the proceedings of Requirements Engineering: Foundation for Software Quality, REFSQ 2015. The chapter also uses material that was published in CTIT Technical Report TR-CTIT-14-05.

In the current iteration, the RASTER method can be performed from beginning to end with reasonable effort, within a lab environment. It is desirable that the method is reliable, that is, that it can be repeated with the same results. Risk assessments methods often have low reliability when they identify risk mitigations for a system based on expert judgement. The goal for this chapter is to assess the reliability of RASTER, and to identify possibilities for improvement of its reliability. This chapter proposes an experimental validation of reliability. A detailed analysis is given of sources of variation, followed by an explanation on how they were controlled and how their mitigations were validated, and a motivation of the statistical procedure used to analyse the outcome. The results can be used to improve the reliability of risk assessment methods, and the approach to validating reliability can be useful for the assessment of the reliability of other methods.

7.1 Introduction

Risks assessments are often performed by experts who assess risks on the basis of best available expert-knowledge of an architectural model. It is known that such expert judgements may have a low reliability [66]. A method is called reliable if it can be repeated with the same results [170]. Other terms in use for this concept are repeatability, stability, consistency, and reproducibility.

Testing the reliability of a risk assessment method is an important issue, which has however received very little attention in the literature. If a risk assessment method is not quite reliable, then its results will always largely depend on the intuition and the expertise of the expert carrying it out. This

weakens the ability of decision makers to justify risk mitigation actions that are based on such assessments.

This chapter illustrates the method that was developed for validating RASTER's reliability. The approach is based on an experiment, guided by a general checklist to ensure that all important aspects are adequately addressed [165]. The sections below illustrate the choices that were made and the methodologies that have been applied to ensure a scientific assessment. The approach is sufficiently general to be applicable to other risk assessment and requirements engineering methods as well.

7.2 Background and related work

To side-step the problem of low reliability of expert judgements, risk assessments are sometimes based on checklists and best practices, in what can be called 'compliance-based methods'. These compliance-based methods are not sufficient for today's telecom networks, mainly because of three reasons. First, telecom operators aim for local optimisations that may have detrimental effects on global availability. One reason for this is that operators may not be aware that their infrastructure is used for the particular service. For example, an operator of fiber-optic transmission network typically leases capacity to other operators and will therefore not know what end-user services are being offered. The end-user organisations' availability requirements are therefore not (fully) known by the operators. Operators strive for high availability and resilience, but are not able to adapt their network to accommodate the availability requirements of individual end users. Second, the infrastructure is extremely complex, and composed of fixed and mobile networks, using PSTN, internet, wireless and cable infrastructures. Third, the infrastructure is in a state of continuous evolution, and hazards to the infrastructure evolve as well. This makes compliance-based risk assessments even less effective than risk-based assessments.

Risk assessment methods can be quantitative (e.g. [41, 123]) or qualitative (e.g. [44, 87]). Quantitative methods estimate probability and impact of risks by ratio scales or interval scales; qualitative methods estimate probability or impact by an ordinal scale, for example 'Low-Medium-High'. Due to lack of information, availability risks for telecom infrastructures have to be estimated qualitatively and by expert judgement. This reduces reliability of risk assessments, either because a single expert makes different estimates for the same object at different times, or because multiple experts make different estimates at the same time. Herrmann et al. argue that reliability is low because risk estimation requires a lot of information [64, 66]. They report that group discussions, although time consuming, have a moderating effect.

Based on related work it can be suspected that RASTER may have low reliability. However, RASTER's group discussions may have a positive effect.

1. Identify and mitigate contextual causes of variation.
Consider causes arising from:
 - (a) Subjects applying the method:
 - i. do they understand the task requested from them?
 - ii. are they able to perform the task?
 - iii. are they willing and motivated to perform the task?
 - (b) Case to which the method is applied.
 - (c) Environment during application.
2. Validate effectiveness of mitigations.

Figure 7.1 – Outline of our method to control contextual causes of variation in reliability experiments (see Section 7.3.1).

7.3 Our approach to testing reliability of a method

Reliability was defined as repeatability, and so it is necessary to know how much variation there is across different instances of using the method, where all possible contextual sources of variation in results have been controlled. It is necessary to understand how to minimise the variation caused by the method itself. Internal causes are inherent to the method, and will be present in any application of the method. For example, the method may be ambiguously described or underspecified. Contextual causes of variation are due to the subject applying the method, the environment during application, or to other aspects of the context in which the method is applied. For example, the time available for application of the method may have been too short. Contextual causes of variation will be present regardless of any particular method being used. A method is considered reliable if the variation of the results of using it is small, when contextual causes of variation are held constant.

The sources of variation in an experiment are different from those in the field. Variation needs to be controlled in order to be able to draw conclusions from the experiment. Controls therefore only need to be effective within the setting of the experiment; it is not necessary that successful mitigations transfer to field settings.

An outline of our approach of keeping contextual causes of variation constant is given in *Figure 7.1*. A description of each of the steps is given below.

7.3.1 Controlling variation

Mitigation of contextual causes of variation involves identification and mitigation of contextual causes, and validation of the effectiveness of mitigations.

Identification and mitigation

Contextual sources of variation can arise from three areas: from the subjects applying the method, from the case to which the method is applied, and from the circumstances and environment in which the method is applied. In practice it is impossible to remove contextual causes altogether, but steps can be taken

to reduce them, or to measure them so that it is possible to reason about their possible influence on the outcome. Because contextual conditions will be controlled, the testing of the reliability of a method will have to be a laboratory experiment (in a field setting, controlling contextual conditions is practically impossible).

Subjects Applying the Method. Three causes were identified for variation arising from the participants in reliability experiments. Participants may not *understand* the task, not *be able* to perform the task, or not *be willing* to do so.

First, *misapplication and misunderstanding of the method by the participants* can cause variation. If the participants do not have a clear understanding of the method and the task at hand, then they may improvise in unpredictable ways. This can be mitigated by providing a clear and concise case which would be easy to explain, together with clear instructions and reference materials. Furthermore, the clarity of these instructions and the task itself can be tested in a try-out, in which the experiment is conducted with a few participants. Experiences from the try-out can then be used to improve the experiment setup.

Second, *lack of experience and expert knowledge* can cause variation. Even when participants understand the method, they still require skills and knowledge to apply the method properly. Researchers in empirical software engineering often use students as subjects in experiments. It is sometimes claimed that the use of students instead of professionals severely limits the validity of software engineering experiments, because students display behaviour that diverges from that of experts. However, it is valid to use students as a model for experts on the condition that students are representative with respect to the properties being investigated. Just like a paper model of a house can be used to study some (but not all) properties of a real house, students can be used to study some (but not all) properties of professionals applying software engineering methods [167]. Some studies have indeed found that certain kinds of behaviour are similar across experts and students [71, 139, 149]. Be this as it may, industry reality often precludes the use of experts in experiments, regardless of how desirable that would be from an experimenter's point of view. Testing with students is cheaper and less risky, and therefore increases the likelihood of successful technology transfer to industry [54]. In addition, in reliability experiments it is not the students' direct results that are of interest. Instead, it is the variation among their results that is relevant. It is therefore not automatically a problem if students achieve different results than professionals, as long as the experiment allows us to draw general conclusions about the method. In the lab (using students) and in the field (using experts) the participants in a reliability experiment should be as similar to each other as possible in background and experience.

Third, participants must *be sufficiently motivated* to apply the method to the best of their abilities. When tired or faced with a tedious and uninteresting task, the quality of the results will suffer. Experiments using students are sometimes

conducted as part of a software engineering course. The experimenter then should consider whether compulsory participation offers sufficient motivation, or whether voluntary participation based on students' intrinsic motivation would be preferable. Furthermore, when the task at hand requires estimation (as will be the case for risk analysis), particular care should be given to avoid personal biases that can result in over- or underestimation. A frequently used way to control this bias is to employ teams instead of individuals. Discussion within the team can dampen individual biases.

Case to Which the Method is Applied. A method such as RASTER is not designed for a single case, but should perform well in a large variety of cases. If a case is ill-defined, then one cannot blame the method if it provides results with low reliability. Since testing of reliability requires a constructed laboratory experiment, the experimenter must carefully design the case to be used by the participants. The case should be representative of cases encountered in practice, but reduced to the essentials to make it fit to the limited time and setting available to laboratory experiments.

Environment During Application. Variation may also derive from *environmental conditions*, such as the meeting room, lighting, or time of day. First, the conditions should be as similar as possible between the different subjects to the experiment. Ideally, conditions should be identical. If the conditions differ, then any variation in results could be attributed to these conditions and does not necessarily indicate a lack of reliability. Secondly, the conditions should be as similar as possible between the experiment and real world applications of the method. For example, the experiment should, or should not be performed under pressure of time, depending on what is the case in practical applications. If the conditions differ, then it could be argued that variation in lab results would not occur in the field.

Validation of effectiveness

When causes of contextual variation have been identified and mitigated, it is necessary to give a convincing argument that mitigation has been effective. The results of the method's application cannot be used in this argument, because the results may vary due to properties of the method rather than due to contextual factors. Instead, it will be necessary to collect additional data, using tools such as interviews, questionnaires, and observations. Therefore experiments to test reliability of a method will collect two kinds of data: measurements on the results of the method, and measurements on the usage of the method. Analysis of these measurements is described next.

7.3.2 Analysis of measurements on the results of the method

The analysis of the reliability of a method can make use of several well-known statistical techniques for inter-rater reliability [57]. Inter-rater reliability is the

Figure 7.2 – Calculating alpha over subsets. D_{expected} , D_{observed}^a and D_{observed}^b must be computed using a common distance metric; D_{expected} and this metric must be based on the relative frequencies of scores over the full data set.

amount of agreement between the scores of different subjects for the same set of items.

Well-known measures for inter-rater reliability are Cohen's kappa, Fleiss' kappa, Spearman's rho, Scott's pi and Krippendorff's alpha [99]. Cohen's kappa and Scott's pi are limited, in that they can only handle two raters. To test reliability, more than two outcomes are necessary in order to be able to draw conclusions. Fleiss' kappa can handle multiple raters but treats all data as nominal. Spearman's rho can take ordinality of data into account, but only works for two raters. Krippendorff's alpha works for any number of raters, and any type of scale. Furthermore, Krippendorff's alpha can accommodate partially incomplete data (e.g. when some raters have not rated some items). This makes Krippendorff's alpha a good choice for this domain. 'Krippendorff's alpha' is abbreviated to 'alpha' in the remainder of this thesis.

Alpha is defined as $1 - D_{\text{observed}} / D_{\text{expected}}$, where D_{observed} is the observed disagreement in the scores and D_{expected} is the expected disagreement if raters assigned their scores randomly. If the raters have perfect agreement, the observed disagreement is 0 and alpha is 1. If the raters' scores are indistinguishable from random scores then $D_{\text{observed}} = D_{\text{expected}}$ and alpha is 0. If $\alpha < 0$, then disagreement is larger than random disagreement. Alpha is therefore a measure for the amount of agreement that cannot be attributed to chance. Cohen's kappa and Scott's pi are basically defined in the same way as alpha, but differ in their computation of observed and expected (dis)agreement. See Appendix A for further details and example computations.

Alpha over subsets

In order to compare the inter-rater reliability of two *subsets* of items, the calculations must be done carefully to ensure that the alphas are comparable. See Figure 7.2. D_{observed} is defined as the average disagreement between scores within units, and D_{expected} as the average disagreement between all scores in the data. For ordinal data, the difference metric (the disagreement between two ordinal values) is based on the relative frequency of scores over all units. The disagreement between two scores will be larger if the ordinal values between them were scored often, and small if the intermediate values were never scored.

When computing alphas over two subsets, the difference metrics used in the calculations must be kept the same for a fair comparison. Also, the D_{expected} value for both alphas must be calculated over the complete set of scores, not over their respective subsets [98]. This computational complexity is not a peculiarity of alpha alone; by analogy, the other measures of inter-rater reliability are affected by a similar issue. Appendix A shows how the regular calculation of alpha must be modified in order to give comparable outcomes over subsets of items, as well as example computations.

7.4 Research method

We conducted a replicated experiment in which small teams of volunteers performed part of the RASTER method on a fictitious telecom service. RASTER is normally performed by teams of telecom and end-user domain experts. However, since experts are scarce and in high demand, it is not possible to rely on expert participation in our laboratory experiment. For practical reasons we opted for a lab experiment instead of a field experiment in a real organisation.

7.4.1 Experiment design

The following description follows the checklist given by Wieringa [165].

Treatment design. Executing RASTER means that several activities need to be performed, ranging from collecting information to obtaining go-ahead from the executive sponsors of the risk assessment (see the overview of the four RASTER stages in *Figure 6.8*). Not all of these are relevant for reliability, because not all of them can contribute to variation in results. Informal case studies suggest that different experts can create architecture diagrams in RASTER that are largely identical. This part of the method was considered reliable, and excluded from the current experiment. In Stage 2, most of the expert assessments of frequencies and impacts of vulnerabilities are made, and so this stage is an important source of possible variation; the case study in Chapter 6 illustrates this. Stages 3 and 4 add no other sources of variation. Including them in the experiment would greatly complicate the experiment without adding new knowledge about sources of variation, and so we decided to restrict the experiment to Stage 2.

Choice of volunteers. To ensure sufficient knowledge of information technology (IT) security, student volunteers were recruited from the Kerckhoffs Masters programme on computer and information security offered jointly by the University of Twente, Eindhoven University of Technology, and Radboud University Nijmegen [96]. Since our groups are not random but self-selected, our experiment is a quasi-experiment [141]. This creates systematic effects on the outcome, that will be discussed in the analysis of the outcomes.

RASTER is applied by a team of experts. This allows for pooling of knowledge and stimulates discussion. Eighteen volunteers were acquired, enabling the

Figure 7.3 – The telecommunication service model used in this experiment: the email service used by a fictitious company. This is the same model as in the handouts in Appendix B.2.

formation of six groups. The sample of groups is not randomly selected and is anyway too small for statistical inference, but similarity-based reasoning is used to draw tentative generalisations to analogous cases [11, 49, 148, 165].

Target of assessment. The telecom service for the experiment had to be small so that the task could be completed in a single afternoon, but large enough to allow for realistic decisions and assessments. The choice of students imposed further restrictions; wireless telecommunication links had to be omitted (as students were unlikely to have sufficient knowledge on these), and a telecom service was chosen to be relatively heavy on information technology. The telecom service for the experiment was an email service for a small fictitious design company heavily dependent on IT systems (Figure 7.3).

Measurement design. For measurement of the results of the method, we used the risk assessment scores by the groups. Groups were instructed to try to reach consensus on their scores. Each assessment was noted on a provided scoring form (one form per group). The possible scores form an ordinal scale. Detailed scoring instructions and descriptions of each of the values were included in the hand-out. In addition, groups could decide to abstain from assessment. Abstentions were allowed when the group could not reach consensus on their score, or when the group members agreed that information was insufficient to make a well-informed assessment.

For measurements on the usage of the method an exit questionnaire and our observations during the experiment were used. Each participant individually completed an exit questionnaire at the end of the experiment. The questionnaire is shown in Appendix B.2.

7.4.2 Using our approach to testing reliability

Subjects Applying the Method. Participants should understand the task, be able to perform the task, and be willing to do so. To mitigate *lack of understanding*, we provided a concise case which would be easy to explain; we prepared what we hoped were clear instructions and reference materials. We then tested the instructions (as well as the task itself) in a try-out. As a result of the try-out, we made small improvements to the instructions and to the case description. At the start of the experiment we invited questions from the participants and made sure to proceed only after all confirmed that they understood the task at hand. To mitigate *lack of experience* (ability to employ RASTER), we created a case that closely matched the expected experience of our students. As explained, we omitted wireless technologies, and emphasised IT systems in the case. To mitigate *lack of motivation* we recruited volunteers, offered the customary compensation, and raffled cinema tickets as a bonus. We stressed the need to avoid personal biases when making estimates.

Case to Which the Method is Applied. Two causes of variation drew our special concern. First, the *number of risk scenarios* could be too large. In the experiment, risks consist of the combination of an architectural component and a vulnerability, e.g. “power failure on the mail server”. Many different scenarios can be devised for this risk to occur. For example, a power cable can be accidentally unplugged, the fans in the power supply unit may wear out and cause overheating, or the server can be switched off by a malicious engineer. A large number of risk scenarios will make the results overly dependent on the groups’ ability to identify all relevant scenarios. Given the limited time available for the experiment, groups could not be expected to identify all possible ways in which a vulnerability could materialise. As a mitigation, we tried to offer clear and limited vulnerabilities in the case description.

Second, reliability cannot be achieved if there is *widespread disagreement on the ‘true’ risk* in society. Physical risk factors can in principle be assessed objectively, but some risk factors (such as fairness or voluntariness) are unavoidably subjective. We therefore use quotation marks for ‘true’; in some cases no single, most valid risk assessment may exist. Such controversial risks do not lend themselves to impartial assessment. In our choice of the experiment’s case we tried to avoid controversial risks.

Environment during Application. We did not identify important causes of variation that needed mitigating. We provided each team with a quiet and comfortable meeting room, in a setting not unlike real world applications of RASTER.

Verification of effectiveness. For each source of contextual variation thus identified, the questionnaire checked whether participants had the required knowledge, ability and motivation to apply the corresponding countermeasure.

For example, for 'lack of knowledge or experience' we used these three questions: "My knowledge of the technology behind office email services can be described as (non-existent – excellent)", "My knowledge of the technology behind office email services could be applied in the exercise", and "It was important that my knowledge of email services was used by the group".

We also used the opportunity to include four questions to test some internal sources of variation. In particular, we wanted to test whether the scales defined for Frequency and Impact were suitable, and whether the procedure to avoid intuitive and potentially biased assessments was effective.

7.5 Results

Each of the six teams scored 138 Frequency assessments and 138 Impact assessments. Our scale for each is (*extremely low, low, moderate, high, extremely high*), but groups were also instructed that they could abstain from scoring. The experiment results can therefore be described as having 6 raters that had to rate 276 items, and partially incomplete, ordinal data. Alpha was computed over these items, but also over subsets of items. Subsets included the Frequency scores and Impacts scores separately, the scores on a single architectural component, and the scores on a single vulnerability.

7.5.1 Scoring results

Over the entire set of items, alpha is 0.338. In the field of content analysis, for which Krippendorff developed his alpha, this is considered a very weak reliability; Krippendorff recommends alpha at least 0.667 for provisional conclusions, and alpha at least 0.8 for firm conclusions, although he stresses that these figures are guidelines only [99]. Over the Frequency scores alpha is 0.232; over the Impact scores alpha is slightly higher, at 0.436.

This relatively low level of agreement is in line with literature about reliability of risk assessments [64, 65]. To understand why the level of agreement is relatively low it is necessary to review the exit questionnaires.

7.5.2 Exit questionnaire

The items of the exit questionnaire are grouped by mitigation measure in *Figure 7.4*; the questionnaire itself contained these items in randomised order (see Appendix B.2). Also, the questionnaire reversed the scales for some of the questions so that 'good' answers do not always belong to the right-most column of the exit questionnaire form. This was also undone in the table shown here. The discussion below also makes use of our observations of the participants during the experiment.

Three causes of variation were mitigated arising from the subjects applying the method: lack of understanding, lack of experience and ability to apply

Lack of understanding

1. The instructions at the start of the exercise were (very unclear – very clear).	VU	Sw Unclr.	N	Sw Clr. 	VC
2. I knew what I needed to do during the exercise.	SD	D	N 	A 	SA
3. In the experiment I could practically apply the instructions that were given at the start of the exercise.	SD	D	N 	A 	SA
4. The instructions that were given at the start of the exercise were (mostly useless – very useful).	MU	Sw Usless.	N	Sw Usefll. 	VU

Lack of experience, ability to apply

5. My knowledge of the technology behind office email services can be described as (non-existent – excellent).	NE	Vry Ltd. 	N 	Good 	Exc.
6. My knowledge of the technology behind office email services could be applied in the exercise.	SD	D	N 	A 	SA
7. It was important that my knowledge of email services was used by the group.	SD	D 	N 	A 	SA

Lack of motivation to avoid biases

8. Before the exercise I was instructed to make rational, calculated estimates.	SD	D 	N 	A 	SA
9. During the experiment I knew how to avoid fast, intuitive estimates.	SD	D	N 	A 	SA
10. The instructions and procedures for avoiding fast, intuitive estimates were (very cumbersome – very easy to use).	VC	Sw C 	N 	Sw EtU 	Vry EtU

High number of risk scenarios

11. When estimating Frequencies and Impacts of vulnerabilities, it is necessary to consider many possible incidents.	SD	D 	N 	A 	SA
12. I could think of practical examples for most of the vulnerabilities.	SD	D	N 	A 	SA
13. When discussing vulnerabilities, other members of my group often gave examples that I would never have thought of.	SD	D 	N 	A 	SA

Controversial risks

14. In my group we mostly had the same ideas on the values of estimates.	SD	D 	N 	A 	SA
15. The estimates made by other groups (compared to ours) will be (very different – very similar).	VD	Sw Dfrnt 	N 	Sw Smlr 	VS
16. For all estimates, there exists a single best value (whether we identified it or not).	SD 	D 	N 	A 	SA

Figure 7.4 – (continued on next page)

Environment

17. I was able to concentrate on the exercise and work comfortably.	SD	D	N 	A 	SA
18. The time to complete the exercise was (way too short – more than sufficient).	WTS	Sw short 	just right 	Sufficient 	MTS
19. Participating in this experiment was (very tiresome – very interesting).	VTs	Sw TS 	N	Interesting 	VI

Suitability of the scales (internal)

20. The scales for values of Frequency and Impact estimates were (very unclear – very clear).	VU	U 	N 	C 	VC
21. In my group we hesitated between two adjacent Frequency and Impact values (almost always – almost never).	AA 	Often 	N 	Seldom 	AN
22. The scales of values for Frequency and Impact were suitable to this exercise.	SD	D 	N 	A 	SA

Rigour of the method (internal)

23. The final answer of my group often equalled my immediate personal estimate.	SD	D 	N 	A 	SA
---	----	-------	-------	-------	--------

Figure 7.4 – Questions and answers to the exit questionnaire; each mark indicates one participant's answer. SD, D, N, A, SA stand for strongly disagree, disagree, neither agree nor disagree, agree, strongly agree respectively. Sw stands for 'Somewhat'. Otherwise, abbreviations are clarified in the text in the first column.

the method, and lack of motivation to avoid personal biases. Answers to the questionnaire (q1–q4) indicate that participants believe they had the required knowledge and instruction to employ the method. Observations during the experiment confirm that, except for a few isolated cases, the instructions were effectively included in the groups' deliberations. The conclusion from this is that our mitigations for lack of understanding were successful.

The answers to q5–q7 (lack of experience and ability to apply the method) were mostly positive, but observations showed a marked difference in practical experience between groups. Some participants, contrary to our expectations, did not fully understand the function and significance of basic IT infrastructure such as DNS servers. To check whether lack of knowledge did induce variation in the scores, the inter-group variation was compared for components that are relatively well-known (such as desktop and laptops), to that for components that are less familiar (such as firewalls and routers). Alphas over end-user components (0.383 for frequencies assessments, 0.448 for impact assessments, 0.416 for the resulting vulnerability levels) were indeed higher than the general scores (0.232, 0.436, 0.338 respectively). The conclusion from this is that lack of experience can explain some of the variation in results.

The answers to q8–q10 suggest that participants succeeded in avoiding personal biases. This was confirmed by the observations. This allows the conclusion that the mitigations for lack of motivation to follow the instructions were successful.

Two causes of variation arising from the case were identified and mitigated: a high number of risk scenarios and widespread disagreement on the 'true' risk. On the number of risk scenarios answers to the questionnaire (q11–q13) and observations indicated that mitigations were successful. In cases when the number of scenarios seemed unlimited (e.g. the risk of a general cable break in the Internet), groups did not hesitate to abstain from answering. For the second cause ("no 'true' risk") the questionnaire results were mixed: positive on agreement within the group and expected agreement with other groups (q14–q15), but negative on whether a single best assessment is possible (q16). The positive results could be a reflection of pleasant, cooperative teamwork, but the negative result to q16 makes it clear that participants believe there is no true answer. Our observations are that most groups made assumptions that significantly affected their assessments. The one group that scored high on q7 ("It was important that my knowledge of email services was used by the group") also was the only group that scored positively on q16. This indicates that the participants probably recognised that their assumptions were somewhat arbitrary. The scoring forms had space for groups to mark important assumptions; none of these assumptions were extraordinary or unrealistic. We did observe that groups generally made many more assumptions than were noted on their forms, but these unrecorded assumptions were mostly natural or obvious. Based on the above, it can be concluded that variation in scores can be partly explained by the difference in assumptions made by groups.

Mitigation of causes of variation from environmental conditions appear to have been successful. Neither questionnaire (q17–q19) nor observations indicate that conditions affected the results unequally. One group finished within the time set for the task, others exceeded that time by a few minutes, although one group finished almost 45 minutes late. All groups completed their tasks.

To summarise, the measurements on the usage of the method indicate two unmitigated contextual causes for variation: participants' lack of experience and knowledge about IT systems, and different assumptions made by the groups. The sources of variation internal to the RASTER method itself are examined next. From these a third cause for variation was discovered.

The questionnaire (q20–q22) and observations showed that groups often hesitated between two adjacent Frequency or Impact classes (recall that all assessments required the selection of a value from an ordinal scale). Participants also remarked that the range of the scales was large, and that the difference between adjacent steps was problematic. We observed that participants volunteered arguments pro and con, and referred to previous scores to ensure a consistent scoring. This was independent of the particular ordinal value; discussion was necessary for the extreme scores as well as for the moderate scores. It is likely that groups settled for different values in these discussions. A third, method-internal cause of variation in outcomes is therefore the difficulty

in choosing between adjacent ordinal values.

7.5.3 Implications

Three explanations were found for the variation in the assessments:

1. The lack of expert knowledge by the participants.
2. The difference in assumptions made by groups.
3. The need to make somewhat arbitrary choices between adjacent ordinal values.

In practical applications of RASTER the team of analysts would consist of industry professionals, and lack of knowledge (1) is therefore not expected. Also, in a field setting analysts have ways to deal with unavailable data other than making assumptions (2). For example, they can make additional observations, conduct inspections, actively look for further documentation, or interview actors to fill gaps in available information. The number and severity of assumptions would therefore likely be much lower in field settings. In practice the team of analysts will be larger than the in the experiment (three students), allowing for more interaction and deliberation in order to reach consensus. Again, this suggests that in practice reliability of RASTER may be higher.

These differences between lab and field suggest that RASTER will produce less variable results in practice than in the lab. Our experiment provides insight in why this can happen; only further field studies can demonstrate whether this is indeed the case.

However, explanation (3) will also be present in the field, and therefore is a point for improvement in the RASTER method.

7.6 Design improvements

From the experiment it was learned that the RASTER method does not effectively avoid variation in its results, because of ambiguity in the definition of classes for Frequency and Impact. From observations during the experiment it can be concluded that there have been no or little difficulties with the interpretation of the extreme and undecided classes (Extremely Low, Extremely High, Unknown, Ambiguous). These were previously called the indefinite classes. Ambiguity is thus mostly limited to the classes Low, Moderate and High, which were defined as the definite classes. Definite classes are those for which experts can possess the knowledge and experience to accurately assess all physical risk factors, given enough time and resources (epistemic uncertainty is low). Indefinite classes are therefore those classes for which epistemic uncertainty make it impossible to accurately assess some or all physical risk factors. The definite classes form an ordinal scale by themselves; two indefinite classes, Extremely Low and Extremely High, can be placed on this scale as well.

<i>Physical factor</i>	<i>Frequency</i>	<i>Impact</i>
Probability of occurrence	•	
Incertitude	•	•
Delayed effects		
Extent of damage		•
Persistency		•
Reversibility		•
Ubiquity/extent of exposure		•

Figure 7.5 – How Frequency and Impact assess the physical risk factors.

For each vulnerability on a component, analysts assess its physical risk factors to obtain its vulnerability level. Our improvement goal is therefore to find a better classification for (groups of) physical risk factors, together with a combination rule that yields risk levels.

Seven physical risk factors were identified while elaborating the requirements in Chapter 4. Currently, Frequency covers the factors Probability of occurrence and Incertitude; Impact covers all remaining factors (except Delayed effects), as well as Incertitude (see Figure 7.5).

Tables 6.5 and 6.6 show the current definitions for the Frequency and Impact aspects respectively. It is not necessary to retain the same classes, nor is it necessary that Frequency has the same kind and number of classes as Impact. However, the number of classes should not be so high that assessment requires more time and effort. The most important goal for redesign is that assessments must have a clear best answer in most of the cases. Analysts should have little trouble determining which of the possible classes has the best match with the risk scenarios that the analysts considered.

It would even be possible to add a new aspect in addition to, or instead of, Frequency and Impact. We considered splitting Impact into two aspects: Damage (covering the physical risk factors Incertitude, Extent of damage, and Ubiquity) and Repairability (Incertitude, Persistency and Reversibility). However, the benefits in clarity did not seem to outweigh the added complexity. Furthermore, the terms Frequency and Impact are familiar to most analysts. We therefore retain the two aspects Frequency and Impact (but probably with redefined classes).

Difficulty to assess Frequency most often derives from difficulty to estimate failure probabilities and to consider failure scenarios, and not so much from ambiguity of the definitions of Low, Moderate and High. The thresholds of these classes currently is at 5, 50, and 500 years respectively. Although it is not known whether these thresholds are the most suitable, it is certain that each of these classes is being scored when applying RASTER, and that the number of High frequencies is outnumbered by Moderate and Low frequencies. There does not seem to be any prior reason to change these thresholds. We therefore focus on redesign of the Impact classes.

<i>Damage</i>	<i>Reversibility</i>	<i>Persistency</i>	<i>Ubiquity</i>
Unnoticeable	Repairable	Short-term	No actors
Degradation	Unrepairable	Long-term	Some actors
Partial unavailability			All actors
Unavailable			

Figure 7.6 – Physical risk factors and their levels used in Impact classes.

<i>Class</i>	<i>Frequency</i>
H High	Partial unavailability, if unrepairable. Total unavailability, if long-term.
M Moderate	Partial unavailability, if repairable. Total unavailability, if short-term.
L Low	Noticeable degradation, repairable or unrepairable.
V Extremely high	Very-long term or unrepairable unavailability of the service.
U Extremely low	Unnoticeable effects, or no actors affected.
A Ambiguous	Indicates lack of consensus between analysts.
X Unknown	Indicates lack of knowledge or data.
– Undetermined	Not yet assessed

Figure 7.7 – Improved Impact classes.

In addition to Incertitude, Impact makes use of four other physical risk factors: Extent of damage, Persistency, Reversibility, and Ubiquity. The levels used for each in the version of RASTER as used in this experiment are shown in Figure 7.6. Reversibility and Persistency can be combined without loss of information into a single factor Repairability with three levels: short-term, long-term, and unrepairable. Using 4 levels for Damage, 3 for repairability, and 3 for Ubiquity, 36 unique combinations can be made (see Figure 7.8). A careful investigation shows that the current Impact definitions do not cover each possible combination, as shown by blank cells in the *Current* column. We propose an alternative that removes the gaps by assigning scores that are reasonable and consistent with existing scores (the *Proposal* column). It turns out that this proposal does not distinguish between cases where some or all actors are affected. Apparently the factor Ubiquity can be disregarded without loss. This fits with the results of the first case study (Appendix B.1) that showed that Delayed effects and Ubiquity are the least relevant to telecommunication availability incidents.

Based on this, the definition of the definite Impact classes was redesigned, as shown in Figure 7.7. The meaning of “short-term” and “long-term” depends on the tasks and use-cases of the actors. A two-minute outage is short-term for fixed telephony but long-term for real-time remote control of drones and robots. “Degradation” means that actors notice reduced performance (e.g. noise during

<i>Damage</i>	<i>Repairability</i>	<i>Ubiquity</i>	<i>Current</i>	<i>Proposal</i>
Unnoticeable	Short-term	No actors	U	U
Unnoticeable	Short-term	Some actors	U	U
Unnoticeable	Short-term	All actors	U	U
Unnoticeable	Long-term	No actors	U	U
Unnoticeable	Long-term	Some actors	U	U
Unnoticeable	Long-term	All actors	U	U
Unnoticeable	Non-repairable	No actors	U	U
Unnoticeable	Non-repairable	Some actors	U	U
Unnoticeable	Non-repairable	All actors	U	U
Degradation	Short-term	No actors	U	U
Degradation	Short-term	Some actors	L	L
Degradation	Short-term	All actors	L	L
Degradation	Long-term	No actors	U	U
Degradation	Long-term	Some actors	L	L
Degradation	Long-term	All actors	L	L
Degradation	Non-repairable	No actors	U	U
Degradation	Non-repairable	Some actors	L	L
Degradation	Non-repairable	All actors	L	L
Partial unavailability	Short-term	No actors	U	U
Partial unavailability	Short-term	Some actors	M	M
Partial unavailability	Short-term	All actors		M
Partial unavailability	Long-term	No actors	U	U
Partial unavailability	Long-term	Some actors	M	M
Partial unavailability	Long-term	All actors		M
Partial unavailability	Non-repairable	No actors	U	U
Partial unavailability	Non-repairable	Some actors		H
Partial unavailability	Non-repairable	All actors		H
Unavailable	Short-term	No actors	U	U
Unavailable	Short-term	Some actors		M
Unavailable	Short-term	All actors		M
Unavailable	Long-term	No actors	U	U
Unavailable	Long-term	Some actors		H
Unavailable	Long-term	All actors	H	H
Unavailable	Non-repairable	No actors	U	U
Unavailable	Non-repairable	Some actors		V
Unavailable	Non-repairable	All actors	V	V

Meaning of letter codes: U stands for extremely low (Unlikely or Ultra low), L for Low, M for Moderate, H for High, V for extremely (Very) high, X for unknown, A for Ambiguous, and - for undecided.

Figure 7.8 – A proposal for improved definition of the ordinal Impact classes. The Current column shows the scores according to the current definitions; blank cells indicate cases not explicitly covered. The Proposal column completes these scores. For meaning of the letter codes see Figure 7.7.

telephone calls, unusual delay in delivery of email messages), but not so much that their tasks or responsibilities are affected. “Partial unavailability or severe degradation” means that actors cannot effectively perform some of their tasks

or responsibilities. For example: email can only be sent within the organisation; noise makes telephone calls almost unintelligible; mobile data is unavailable but mobile calls and SMS are not affected. Actors can still perform some of their tasks, but other tasks are impossible or require additional effort. "Total unavailability" means that actors effectively cannot perform any of their tasks and responsibilities using the telecom service. When the impact is extremely high, major redesign of the telecom service is necessary, or the service has to be terminated and replaced with an alternative.

7.7 Conclusion

From these results some general conclusions can be drawn.

The previous sections present an approach to validating and measuring the reliability of a method, and present a research design that uses this approach. Our approach to measuring reliability of methods does not mention risk assessment at all, and should be of use also in measuring the reliability of other methods. The research design too should be of use for measuring other properties of methods.

Statistical tools for measuring inter-rater agreement can be used to measure the reliability of methods that rely on experts for qualitative or quantitative assessments. For subsequent analysis of reliability, it is often useful to compare inter-rater agreement on certain subsets of scores. Standard statistical measures need to be applied carefully and modified appropriately in order to make comparisons possible. We have shown how this should be done for one statistical measure, namely Krippendorff's alpha. For other measures the modifications are similar.

The analysis confirms the findings from literature that reliability of expert judgements of likelihood and impact is low. Our results add to this a quantification of that lack of reliability, using statistical tools, and a careful explanation of all possible sources of this lack of reliability, in the method as well as in its context of use. From our quantitative analysis it can be concluded that reliability in risk assessment with scarce data, that has to rely on expert judgement, may not be able to reach the standards common in content analysis. It may very well be the case that it is not achievable for any method that requires a very high amount of expert knowledge.

If this is true, then experts performing such assessments retain a large responsibility for their results. Their risk assessments not only yield risk evaluations, but also the limitations on justifications for these evaluations, along with best guesses of likelihood and impact. They have to communicate these limitations of their evaluations to decision-makers who use their evaluations.

The improvements made as a result of this experiment complete the design

of the method. Other than updates to the tool, the final version of the RASTER method is shown in Appendix D. This is the version that will be validated in two field tests, as described in the next two chapters.

First Field Test

This chapter is an expanded version of the author's publication "Validating the Raster Risk Assessment Method in Practice" in the proceedings of the 12th International Conference on Information Systems for Crisis Response and Management ISCRAM2015.

The next step in this research is to validate the improved method in practice, under uncontrolled circumstances. This step is to answer whether, in the field, RASTER yields correct results against an acceptable cost, and whether experts perceive the method as usable and useful. For these field tests three Safety Regions and a Water Board were approached. Eventually two host organisations were found that were both amenable to the need for risk assessment and able to participate in the test. This chapter describes the field test.

The first field test was hosted by Safety Region Groningen (SRG). The project was expedited through previous contacts between SRG and Radiocommunications Agency Netherlands in its role as a government agency. Several informal and formal meetings with the management of SRG were held to obtain approval for the project. Subsequently, five work sessions with experts from the crisis organisation were needed to complete all the risk assessments. The projects' basic results, if printed, would be about seventy pages. A thirty-page internal report summarised the assessments, describing the top-priority availability risks and recommendations for risk treatment. This report is confidential, because it explicitly describes weaknesses in telecommunication systems. The results were reported to management, and most treatment recommendations have since been implemented.

8.1 Setting and method

As explained in Section 1.2.2, crisis management in the Netherlands is organised in Safety Regions. Our object of study in this field test is the Safety Region Groningen (SRG). SRG serves a relatively sparsely populated coastal province

of about 2,300 km² and a population of 575 thousand people. The region has a single large city, and is characterised by installations for the winning of natural gas. As from 2014, SRG has an information management department that is responsible for information and communication systems deployed by the Safety Region. This department was the sponsor of the study. Their objective was to receive an independent assessment of whether the level of reliability of their services was suitable, given the needs of its internal customers.

The research objective for this project is to validate the RASTER method, as described above. The design of our validation follows the checklist provided by Wieringa [165]. In Wieringa's terminology, the field test described in this chapter is *technical action research*, an application of a still-experimental technique in the real world to help a client. In the field test described here, the author (the experimenter) participated in a project within SRG in a real operational environment. The experimenter acted as project leader, facilitating the application of the RASTER method. The goal of the field test was therefore twofold: the sponsor (as the client) had an actual question that needed answering, and the experimenter wanted to validate the RASTER method.

8.1.1 Research questions

For validation three research questions need answering:

- Q1 Does RASTER produce, in practice, risk evaluations that are correct? That is, are all relevant risks included (completeness), are low risks excluded (conciseness), and are risks presented with the right priorities?
- Q2 What are the costs and effort involved in execution of RASTER?
- Q3 Are the target users willing to use the RASTER method in their practice?

Each question is described in turn.

Q1: Correctness. Correctness of risk evaluations is a difficult concept. First, it cannot be determined objectively, because there is no available known good standard to compare against. Secondly, information is incomplete and the required amount of expert judgment is necessarily large. Lastly, risk evaluations are uncertain predictions, meaning that they state that some future impact is likely or less likely. Crises are infrequent, but even when the impact materialises the fact that the event did happen in no way validates or invalidates the risk assessment. It is only with a large set of predictions and their outcomes that statistical statements can be made about the accuracy of predictions. In that respect risk evaluations for crises are very different from other uncertain predictions, such as weather forecasts. Because of these difficulties, correctness must be determined in a subjective way. Firstly, we asked participants about their personal belief in correctness of their risk evaluations; secondly, we asked participants from an independent field test to assess the correctness of these evaluations. This cross-validation is described together with the second field

test, in the next chapter.

Q2: Required effort. Since the parties agreed to participate without charge, only time and effort were relevant. We recorded the time spent during project meetings, and the participants reported the amount of time they spent on project matters between meetings.

Q3: Disposition to use. To see whether target users would want to use RASTER themselves, we used the Technology Acceptance Model TAM [31]. TAM has been studied extensively and, although it has received criticism and several extensions have been proposed, its usefulness is well established in research. We use TAM because it employs a list of 12 standardised questions that accurately predict whether users would adopt new information technology. The list was integrated easily into our own post-test questionnaire (see the next section). The RASTER method falls within the scope of TAM, because the use of the RASTER software tool forms an essential part of the method.

TAM measures perceived usefulness and perceived ease of use as predictors for future behaviour; TAM does not (nor does it intend to) measure objective benefits or efficiency gains. The standard TAM questions were translated from English into Dutch.

8.1.2 Method

In Section 3.2.2 the continuity plan framework used by Dutch Safety Regions was described. The SRG adopted the framework and completed a continuity plan based on it. The main author of this plan did not participate in the field test and did not influence its results but the contents of the plan were available during the project for reference.

Before the project was started, a project plan was presented to the sponsor for approval. The plan presupposed the participation of experts from diverse backgrounds. These experts would perform the RASTER method, with the experimenter acting only as facilitator and chair. After receiving approval on the project plan a kick-off meeting was held, at which participants from the various disciplines were present. In addition to the initial project members (see list in *Figure 8.1*) there were representatives from the management team of SRG. The managing director of SRG was present to express his support.

At the end of the project, each participant completed a questionnaire, and a guided group discussion was held. The purpose of the questionnaire and discussion was to provide further information to answer our three research questions. The questionnaire contained 38 statements with five possible answers from 'Strongly disagree' to 'Strongly agree'. Together with the 12 TAM questions this yielded 50 questions in total; Appendix B.3 shows the questionnaire.

Given the sensitivity of SRG's operations, all participants agreed to confidentiality. Publishing details of the risk analysis or its results could jeopardise the effective operation of emergency services, or could make them a target for

<i>Organisation</i>	<i>Role and domain expertise</i>	<i>Member</i>
emergency medical care	ambulance transport and medical emergency response	yes
fire services	professional, emergency response for fires and accidents	no
fire services	volunteer, emergency response for fires and accidents	yes
municipalities	civil support and crisis communication	no
police	public safety, law enforcement and decision support	yes
SRG	information management and telecommunications	yes, core
SRG	information management and telecommunications	yes, core
water board	information technology	yes
water board	water management and decision support	yes
experimenter	RASTER method and meeting chair	—

Figure 8.1 – List of initial participants to the field test: their affiliation, role within the Safety Region, and their particular expertise. Participants who dropped out are marked ‘no’; others are (core) project members.

malicious actions. This chapter therefore only describes the results in general terms and does not mention specific risks found.

8.2 Results

All project members were present (with a few apologies) during each of the five half-day project meetings. Participants who dropped out before or just after the first project meeting are not counted as project members. Using this definition, there were eight project members, including the experimenter (*Figure 8.1*). The purposes of plenary meetings were to introduce and explain the steps of the RASTER method, to collect ideas from all project members, and to collectively discuss the main approach to conducting the risk analyses. The experimenter and members from the information management department (the *core* members in *Figure 8.1*) convened between plenary meetings to complete tasks that were left unfinished during the plenary meetings. The full results were then presented at the next plenary meeting for the group’s approval. This way of working allowed for efficient use of the expert’s time.

The group inspected ten telecommunication services: *Figure 8.3* provides an overview of the complexity of their diagrams and *Figure 8.2* shows an anonymised version of the Voice over IP service. Together these ten services contained 205 telecom infrastructure components (157 when excluding actor nodes, which do not contain risk assessments). In all, the project recorded over 1,800 estimates of likelihood and impact of vulnerabilities (see *Figure 8.4*). The final risk list contained 19 important availability risks and their treatment recommendations. Details on the time spent by the project team are given in Section 8.3.1.

Figure 8.2 – The project’s diagram for fixed telephony (anonymised). Components have a shape, colour (shown in greys here) and emblem indicating type, location and risk level respectively. This is the second most complex diagram (check Figure 8.3).

Exit questionnaire

At the project’s end seven participants completed the exit questionnaire (Figure 8.5). A summary of the questionnaire results is given below; Appendix B.3 compares its results to those of the second field test. Participants were able to participate in the project (questions 1–4), able to contribute to its results (questions 5–7), and had little difficulty resolving disagreements and attaining consensus (questions 8–9). The scales for estimates of likelihood and impact were clear and suitable to the project (questions 10–16). The time available for the project was somewhat tight, but sufficient (questions 17–20), and participants happily made themselves available (questions 21–22). Lack of documentation in Dutch was perceived as an issue in general, although not all participants were personally affected by this (question 23–24). Participants agreed that RASTER helps to weed out small risks, but less certain whether it was able to capture all large risks (question 25–28). Still, they believe the outcome of the project to be right (question 29–32) although not very effective (question 33–34). The process of creating diagrams and discussing vulnerabilities of components was found very useful (question 35–38).

Group discussion

Afterwards, a one-hour guided discussion was held. The topics for this discussion were: 1) the RASTER methods and its strengths and weaknesses, 2) reliability of the outcome, 3) time and effort, and 4) determination of correctness. Unless

<i>Service</i>	<i>Description</i>	<i>Wireless</i>	<i>Wired</i>	<i>Equip't</i>	<i>Cloud</i>	<i>Actor</i>	<i>Total</i>
data services	net-centric operations	7	7	30	15	9	68
fixed telephony (over IP)	external and general communications	1	5	11	10	6	33
civil warning sirens	to warn the public about crises or disasters	8	0	11	3	5	27
vehicle automation	on-board navigation and special tools	1	1	8	11	3	24
private mobile ra- dios	for group communication and despatch	2	2	9	4	6	23
paging	group and personnel notification	1	0	9	4	5	19
mobile telephony	external and general communications	0	0	5	6	4	15
VPN connections	inter-office communications	0	4	6	3	2	15
national emergency telephony service	last resort communications	0	2	7	4	4	17
satellite telephony	backup voice communication	1	2	5	2	4	14
Total:		18	20	93	26	48	205

Figure 8.3 – Telecom services analysed in this field test, and the number of nodes for each node type. Not all columns add up to the provided totals because some nodes appear in more than one diagram.

Figure 8.4 – Summary of the first field test: number of nodes per type, and number of risks per level (Stage 2, single failures only).

indicated otherwise, the opinions stated in the summary of the comments below are the participants’.

1) Although the results of the project were not highly surprising, none of the participants believed that they were led towards the expected outcome. They remarked that the structured method allowed for an unbiased procedure, the results of which could not always be predicted in advance. One of the most important results is that the participants gained a better insight in the use of telecom services and their infrastructure. As an illustration, one participant

Ability to participate

1. The explanations and instructions on the RASTER method was (very unclear ... very clear).	VU	U	N	C ●●●●●	VC ●●
2. The explanations and instructions on the goals of the project were (very unclear ... very clear).	VU	U	N	C ●●●●●	VC ●
3. I knew what input was expected from me during the project.	SD	D	N ●	A ●●●●●	SA
4. I was able to apply the explanations and instructions on the RASTER method.	SD	D	N ●●	A ●●●●●	SA

Ability to contribute

5. My technical knowledge on telecom services is (very limited ... excellent).	VL	Limtd. ●●	N ●●	Good ●●●	Exc.
6. My knowledge on the use and purpose of telecom services in practice is (very limited ... excellent).	VL	Limtd. ●	N ●●	Good ●●●●	Exc.
7. With my knowledge I was able to contribute to the project.	SD	D	N ●●	A ●●●●	SA ●

Attitude to consensus

8. In the project group we agreed on the estimates for Frequency and Impact.	SD	D	N ●	A ●●●●●	SA ●
9. In principle there is always one best estimate (whether we always found it or not).	SD	D	N ●●●●●	A	SA ●

Clarity of the scales

10. The scale used for Frequency was (very unclear ... very clear).	VU	U	N	C ●●●●●●	VC
11. The scale used for Impact was (very unclear ... very clear).	VU	U	N	C ●●●●●●	VC

Applicability of the scales

12. When making estimates on Impacts I ... hesitated between two neighbouring classes (almost always ... very seldom).	AA	Often ●	N ●●	Seldom ●●●●	VS
13. When making estimates on Frequencies I ... hesitated between two neighbouring classes (almost always ... very seldom).	AA	Often ●	N	Seldom ●●●●●	VS
14. The scale used for Frequency was suitable for this project.	SD	D ●	N	A ●●●●●	SA
15. The scale used for Impact was suitable for this project.	SD	D	N	A ●●●●●●	SA
16. The group's estimated matched my personal estimates.	SD	D	N ●●●	A ●●●	SA ●

Figure 8.5 – (continued on next page)

Time required by the analysts

17. The time available for this project was (very tight ... more than ample).	VT	Tight ●●	N ●●●●	Ample	MTA
18. If we had spent more time on the project, the results would have been (much worse ... much better).	MW	W	N ●●●●●	B ●	MB
19. It was easy to make time available for this project.	SD	D ●●●●	N ●	A ●●	SA
20. If the project had required more time, I would not have been able to participate.	SA	A ●	N ●●●	D ●●●	SD
21. The project took a lot of my time.	SA	A	N ●●●●	D ●●●	SD

Attitude to the project

22. Participating in this project was (very boring ... very interesting).	VB	B	N	I ●●●	VI ●●●●
---	----	---	---	----------	------------

Need for localisation

23. I found it difficult that the RASTER manual was only available in English.	SD	D ●●	N ●●	A ●●●	SA
24. It is important that all RASTER documentation is available in Dutch.	SD	D ●	N	A ●●●●●	SA ●

Accuracy and efficiency of risk identification

25. The RASTER method assists in quickly determining all large risks.	SD	D ●	N	A ●●●●●	SA ●
26. There exist large risks for unavailability of telecom services that were not found in this project.	SA	A ●	N ●●●●	D ●	SD ●
27. The RASTER method helps to ignore all small risks quickly.	SD	D ●	N	A ●●●●●●	SA
28. The final risk list also contains risks that actually are not that important.	SA	A ●	N ●	D ●●●●	SD ●

Correctness of the outcome

29. I am confident that the results of the project are correct.	SD	D	N ●	A ●●●●	SA ●●
30. I can take responsibility for the results of the project towards my colleagues.	SD	D	N	A ●●●●	SA ●●●

Correctness of risk prioritisation

31. The RASTER method helps to determine risk priorities.	SD ●	D	N	A ●●●●	SA ●●
32. The risks on the final risk list are shown in the right order (highest priority first).	SD	D	N	A ●●●●●●●	SA

Effectiveness of the project

33. I am surprised by the results of this project.	SD ●	D ●●	N ●●●	A ●	SA
34. Altogether, the project has made [organisation] (much less secure ... much more secure).	MLS	Less S.	N ●●●●	More S. ●●●	MMS

Figure 8.5 – (continued on next page)

Clarity of procedures

35. For me drawing and discussing the telecom diagrams was (very unnecessary ... very useful).	VU	Somewhat Unnry.	N	Somewhat Usefl. ●	VU ●●●●●
36. I found the telecom diagrams (very unclear ... very clear).	VU	U	N ●	C ●●●	VC ●●●
37. For me estimating and discussing Frequency and Impact was (very unnecessary ... very useful).	VU	Somewhat Unnry.	N ●●	Somewhat Usefl.	VU ●●●●●
38. I found the estimates of Frequency and Impact (very unclear ... very clear).	VU	U	N ●	C ●●●●●	VC

Technology Acceptance Model – Perceived Ease of Use

39. Learning to operate RASTER would be easy for me.	SD	D	N ●	A ●●●●●	SA
40. I would find it easy to get RASTER to do what I want it to do.	SD	D ●	N ●●	A ●●●	SA
41. My interaction with RASTER would be clear and understandable.	SD	D ●	N ●	A ●●●●●	SA
42. I would find RASTER to be flexible to interact with.	SD	D	N ●●●●	A ●●	SA ●
43. It would be easy for me to become skilful at using RASTER.	SD	D ●	N	A ●●●●●	SA
44. I would find RASTER easy to use.	SD	D	N ●●●	A ●●●●	SA

Technology Acceptance Model – Perceived Usefulness

45. Using RASTER in my job would enable me to accomplish tasks more quickly.	SD	D ●●	N	A ●●●●	SA ●
46. Using RASTER would improve my job performance.	SD	D	N	A ●●●●	SA ●●●
47. Using RASTER in my job would increase my productivity.	SD	D ●	N ●●●	A ●●	SA ●
48. Using RASTER would enhance my effectiveness on the job.	SD	D	N ●●	A ●●●●●	SA
49. Using RASTER would make it easier to do my job.	SD	D	N	A ●●●●●	SA ●
50. I would find RASTER useful in my job.	SD	D	N ●●	A ●●●●	SA ●

Figure 8.5 – Questions and answers to the exit questionnaire for the first field test; each mark indicates one participant's answer. SD, D, N, A, SA stand for strongly disagree, disagree, neither agree nor disagree, agree, strongly agree respectively.

assumed that some other party had satellite phones available for operational use, but this turned out to be not the case. This insight allows participants to better appreciate the merit of risk reduction measures. For example, it may not be useful to have redundant connections using separate geographical routers if these connections terminate within the same wiring closet. The diagrams were found to be particularly useful; the use of colour and risk level emblems makes that diagrams quickly and easily convey useful information. They often referred

to the definitions of the different classes for Frequency and Impact. Using classes instead of making precise estimates was seen as a time-saver; it would not have been constructive to, for example, discuss whether a vulnerability would occur once every five or ten years. The inclusion of social risk factors was seen to be a useful addition. It made that some risks were prioritised differently in the Evaluation stage, as citizens have a different view on risks than insiders.

2) With different participants the outcome would have been roughly the same, provided that the same time and effort would be spent. Participants also noted that not having a technical background was not an obstacle to contributing to the team. If the project were redone, it would be recommendable to include a participant from the common emergency call centre and dispatch. Other than the call centre organisation, no other organisations were missing from the project team.

3) If more time had been available, then probably extra time would have been spent on checking of assumptions, and on interviews of the providers of telecom services. The results would have been more accurate, but not necessarily more useful. There is always a trade-off between level of detail and practical applicability of the results. Reducing the lead time by compressing the time between work sessions would not have been productive. Participants stated that time is needed for reflection, as well as for making queries to colleagues and suppliers.

4) Participants agreed that estimates are partly subjective. Members participate because of their expertise, and it is their responsibility to make the best possible estimates and judgements. Because of the team composition, subjectivity will be averaged out. The structure of the RASTER method also helps in avoiding subjectivity in the outcome.

8.3 Discussion

All participants were provided with a booklet explaining, in English, the RASTER method and tool. It surprised us that experts found the lack of a Dutch translation a stumbling block (question 23, 24). A few participants even preferred to use an online translation tool over the English version.

The project again confirmed that tool support is essential. In the group discussion, participants indicated that they found the use of colour in diagrams very useful to convey relevant information. (Unfortunately, all the examples in this thesis are in greys only).

Overall, the participants found the project interesting, stimulating, and very useful, not just because of the final risk list but also because it gave them an understanding of the workings and complexities of telecom services that otherwise remain hidden behind the scenes. This may translate to better preparedness and effectiveness; the results do not allow a conclusion to be

drawn about that.

Finally, an interesting question is how the risk list from this project compares to the results from the continuity plan framework. Comparing the two is difficult, because the information management department of SRG had made significant changes to the infrastructure after the continuity plan was written, and before the RASTER assessment was made. For example, all servers had been outsourced and desktop equipment currently consists of thin clients; fixed telephony had been replaced by a hosted VoIP solution. Nevertheless, we observed that some of the telecom services that contained the most risks in the RASTER project (which, for reasons of confidentiality, cannot be named) were assessed as “reliable” in the continuity plan. It can easily be argued that RASTER gave the more accurate description, because its risk evaluations are grounded in a detailed analysis of the technical architecture of the telecom services, whereas the continuity plan did not offer any argumentation for its assessments.

The project ran smoothly and easily, project members found the risk analysis to be interesting and participated in lively discussions. Using the coloured RASTER diagrams as a common ground, responders and planners were able to collaborate with telecom experts. Together they drafted and justified risk evaluations that were approved by all participants as well as the sponsor.

The results also allow the tentative conclusion that RASTER’s results are complete, concise and prioritised correctly. The effort required, about 140 staff-hours over a period of six weeks for a medium-sized crisis organisation, is significant but appears to be acceptable. Perceived ease of use and perceived helpfulness suggest that the RASTER method would be adopted by its target users, provided that localised versions of the manual are available. Successful implementation will probably require localised training materials, such as examples and tutorials, as well.

These results suggest to us that RASTER is feasible and useful in practice.

8.3.1 Research questions

This subsection answers the three research questions.

Q1: Correctness. As for completeness, almost all participants agreed that RASTER helps to find all large risks quickly (86% agreed or strongly agreed to question 25), but were less certain when asked whether there are large risks that have not been found (14% agreed, 57% had no clear opinion, question 26). For conciseness, participants almost all agreed that RASTER helps to ignore small risks quickly (86% (strongly) agreed, question 27), but were slightly less certain whether the final risk list contains risks that are actually not that important (71% believed this not to be the case, question 28). All participants agreed that the priorities on the risk list were right (question 32). We also asked whether participants trusted the results of the project (86% agreed strongly, question 29), and whether they were willing to take responsibility for the results towards

Figure 8.6 – Scores on the TAM questions on Perceived Ease of Use.

their colleagues (all agreed, question 30). Overall, it can be concluded that participants believe the results to be correct, but also that they recognise that uncertainties are still large. This corroborates the findings in Chapter 7 about generally low reliability of experts' assessments of risk.

Q2: Required effort. The second research question concerns the effort required in execution of RASTER. Five project members spent on average 16 hours on the entire project; the three core members who actively worked on the project between meetings spent on average 29 hours. The project members spent 26 hours, 60 hours, 6 hours, and 21 hours respectively on the four stages of the RASTER method, and 24 hours on the kick-off and evaluation (see Figure 9.6). Their combined total of 137 staff-hours excludes time spent by the experimenter. According to the participants, the effort was sufficient (question 18): only one participant believed that the results would have been better if more time had been available, all others scored neutral. Also, it would not have been impossible for them to participate if the required time had been higher (question 20).

Q3: Disposition to use. The questionnaire included the 12 TAM questions in random order and position. The questions on perceived ease of use (Figure 8.6) were answered positively, and the answers for perceived usefulness (Figure 8.7) were even higher. These are positive results, especially since perceived usefulness is an even stronger predictor of future use than perceived ease of use [31].

8.4 Design improvements

The project made clear that documentation and tools need to be provided in the native language of experts. We have therefore localised the tool, and translated the RASTER manual; both are now available in English and Dutch.

We also found that this project contained many more components than in our previous lab experiments. The user interface of the RASTER tool for editing

Figure 8.7 – Scores on the TAM questions on Perceived Usefulness.

clusters for analysis of common cause failures turned out to be impractical for this large number of components. We improved the tool, so that multiple components can be selected quickly and assigned to a cluster in a single action, either by dragging them as a group or through a popup menu. This example shows that the prototyping aspects of the tool are indeed useful in practice.

8.5 Conclusion

From these results some general conclusions can be drawn.

Management of crisis organisations are well aware of the need to perform risk assessments in general, and for unavailability of telecommunication services in particular. They are acutely aware of the consequences should telecom services fail, but depend on their staff for technical expertise and risk assessment. A project proposal with clearly articulated goals and planning is necessary, but given a well-motivated plan it is possible to claim capacity for seven experts to convene for several half-day risk assessment sessions, even though expert's time is scarce and expensive.

When experts from technical and non-technical domains perform a group-wise risk assessment a rich graphical model can facilitate collaboration. The use of colours, shapes and notes enrich graphical models, and make reasoning about architectures easier for non-technical experts. This results in better arguments for the risk assessments obtained.

Reliability of expert risk assessments (of likelihood and impact) is generally low. Compared to the reliability experiment in Chapter 7, the participants to the field test had relevant domain knowledge and ample working experience. Still, in their opinion the results are subjective and retain uncertainty; they are prepared to take personal responsibility for their choices, assessments, and recommendations.

Second Field Test

This chapter has not yet been published, but a Technical Report is available, and an overview paper describing the design and validation of the entire RASTER method will be submitted to a journal.

We performed a second field test at Water Board Hunze en Aa's (WHA), a water board in the north-east of the Netherlands. WHA is an active participant in the Safety Region Groningen, and it was the liaison for the water board participating in the first field test at the safety region who took the initiative to perform a similar project at WHA.

9.1 Setting and method

WHA is a public authority responsible for water management in a north-eastern region of the Netherlands. Its task is to ensure adequate water levels and to maintain water quality standards. Specifically, WHA maintains dikes and other water barriers, regulates the ground and surface water levels, maintains waterways used for shipping, and manages waste water treatment installations. Nature, agriculture, recreation and physical safety all pose their specific requirements on water management, and WHA takes the requirements of these stakeholders into account. WHA actively participates in the operational activities of the Safety Region Groningen, and to a lesser extent the Safety Region Drenthe.

A risk assessment reveals vulnerabilities and company confidential information. The full results of the assessment therefore have to remain confidential, as in the first project. This chapter can only describe the results in general terms, and does not discuss specific vulnerabilities.

Within WHA, its IT department is responsible for both information technology and communication facilities. This includes office automation, telecommunication connections to more than a dozen external locations (water treatment plants, pumping stations, locks, labs and maintenance shops), and telemetry connections to hundreds of objects. Technologies used include fiber-optic cables,

<i>Role</i>	<i>Domain expertise</i>
sponsor	crisis management
member	information technology and data communications
member	public enforcement of regulations
member	telemetry
member	water level management
experimenter	RASTER method and meeting chair

Figure 9.1 – List of participants: their project role and their particular expertise.

DSL links, point-to-point microwave radios, and public mobile data networks based on 2G, 3G and increasingly 4G technologies.

Early in 2015 the IT department awarded a tender for telecommunication services. In the course of 2015 and 2016 services will be migrated and updated, and during this process the infrastructure will contain a complex mixture of ‘old’ and ‘new’ services. The main goal of WHA in participating in this field test is to mitigate existing availability risks in the current infrastructure, and to avoid introducing new availability risks.

The scope of the RASTER project of WHA contained those telecommunication services for which WHA (specifically its IT department) had technical or functional responsibility. As always, the first stage of the project defines precisely which telecommunication services are within scope.

Method

We used the same research questions and research method as for the first field test (see Section 8.1). This means we have three research questions, use Technical Action Research, collect an exit questionnaire identical to the one used before (Appendix B.3), and guide an evaluation discussion.

After discussions with the project sponsor on the project scope and planning, a series of project meetings was started with various professionals from WHA. The sponsor joined the project, representing the crisis management team. Three experts from the IT, telemetry, and enforcement departments participated; later a fourth expert from water level management complemented the project team (see Figure 9.1).

In this field test, in addition, we performed two comparisons between the field tests: we compared and contrasted the outcomes, and also asked the participants to reflect on the outcomes of the first field test (to perform a limited peer review). The goal of this review is to validate reliability (repeatability) under practical circumstances, and to collect additional data on the correctness of the results. The review is described in Section 9.4.

Figure 9.2 – The project’s diagram for telemetry (anonymised). Components have a shape, colour (shown in greys here) and emblem to indicating type, location and risk level respectively.

9.2 Results

In total ten project meetings were held (including one evaluation session), lasting three to four hours. Most project members were present during the meetings, with the exceptions of meetings six and eight, when only two experts were present. Project members convened only incidentally between meetings. Because the project took longer than planned, a three week break over the summer holidays had to be included. Towards the finalisation of the project there was also a few weeks idle time because of scheduling reasons.

The project covered seven telecommunication services: *Figure 9.3* provides an overview of the complexity of their diagrams and *Figure 9.2* shows an example of the fixed telephony service. Together, these seven services contained 204 telecom infrastructure components. In all, the project recorded just over 1,900 estimates of likelihood and impact of vulnerabilities (see *Figure 9.4*). The final risk list contained nine important availability risks.

<i>Service</i>	<i>Description</i>	<i>Wireless</i>	<i>Wired</i>	<i>Equip't</i>	<i>Cloud</i>	<i>Actor</i>	<i>Total</i>
inter-office	wired and wireless connections between various offices, work locations and objects	5	10	22	24	3	64
telemetry	remote monitoring of water levels and objects	7	12	17	11	7	54
national emergency telephony service	last resort communication	0	9	10	4	6	29
fixed telephony	external and general communications	0	8	6	5	7	26
mobile telephony	external and general communications	4	0	6	5	7	21
private mobile radios	group communication during crises	1	4	8	2	3	18
marine radio	ship-to-ship and ship-to-shore	1	0	3	0	3	7
Total:		18	43	72	36	35	204

Figure 9.3 – Telecom services analysed in this field test, and the number of nodes for each node type. Not all columns add up to the provided totals because some nodes appear in more than one diagram.

Figure 9.4 – Summary of the second field test: number of nodes per type, and number of risks per level (single failures only).

Exit questionnaire

All five participants completed the exit questionnaire; the questionnaire was identical to the one used in the first field test. One question was left unanswered: because the RASTER manual had since been translated into Dutch, the statement “I found it difficult that the RASTER manual was only available in English” was no longer relevant. A summary of the questionnaire results is given below; Appendix B.3 compares its results to those of the first field test.

Participants found the instructions to be clear, but some had difficulty in applying them to their contribution in the project (questions 1–4). They were able to contribute to its results (questions 5–7), and had little difficulty resolving disagreements and attaining consensus (questions 8–9). Although the scales

Ability to participate

1. The explanations and instructions on the RASTER method was (very unclear ... very clear).	VU	U	N	C ▲▲▲▲	VC ▲
2. The explanations and instructions on the goals of the project were (very unclear ... very clear).	VU	U	N ▲	C ▲▲▲	VC ▲
3. I knew what input was expected from me during the project.	SD	D ▲	N ▲	A ▲▲	SA ▲
4. I was able to apply the explanations and instructions on the RASTER method.	SD	D	N ▲▲	A ▲▲▲	SA

Ability to contribute

5. My technical knowledge on telecom services is (very limited ... excellent).	VL	Limtd. ▲▲	N ▲▲	Good ▲	Exc.
6. My knowledge on the use and purpose of telecom services in practice is (very limited ... excellent).	VL	Limtd.	N ▲▲▲	Good ▲▲	Exc.
7. With my knowledge I was able to contribute to the project.	SD	D	N ▲▲	A ▲▲▲	SA

Attitude to consensus

8. In the project group we agreed on the estimates for Frequency and Impact.	SD	D	N ▲	A ▲▲▲	SA ▲
9. In principle there is always one best estimate (whether we always found it or not).	SD	D	N ▲▲▲	A ▲▲	SA

Clarity of the scales

10. The scale used for Frequency was (very unclear ... very clear).	VU	U	N ▲	C ▲▲▲▲	VC
11. The scale used for Impact was (very unclear ... very clear).	VU	U	N	C ▲▲▲▲▲	VC

Applicability of the scales

12. When making estimates on Impacts I ... hesitated between two neighbouring classes (almost always ... very seldom).	AA	Often ▲▲▲▲	N ▲	Seldom	VS
13. When making estimates on Frequencies I ... hesitated between two neighbouring classes (almost always ... very seldom).	AA	Often ▲▲	N ▲▲▲	Seldom	VS
14. The scale used for Frequency was suitable for this project.	SD	D ▲▲	N	A ▲▲▲	SA
15. The scale used for Impact was suitable for this project.	SD	D ▲	N ▲	A ▲▲	SA
16. The group's estimated matched my personal estimates.	SD	D ▲	N ▲▲	A ▲▲	SA

Figure 9.5 – (continued on next page)

Time required by the analysts

17. The time available for this project was (very tight . . . more than ample).	VT	Tight	N ▲▲▲▲	Ample ▲	MTA
18. If we had spent more time on the project, the results would have been (much worse . . . much better).	MW	W	N ▲▲▲▲	B ▲	MB
19. It was easy to make time available for this project.	SD	D ▲▲▲	N ▲	A ▲	SA
20. If the project had required more time, I would not have been able to participate.	SA ▲	A	N	D ▲▲▲	SD ▲
21. The project took a lot of my time.	SA	A	N ▲▲▲	D ▲▲	SD

Attitude to the project

22. Participating in this project was (very boring . . . very interesting).	VB	B	N	I ▲▲	VI ▲▲▲
---	----	---	---	---------	-----------

Need for localisation

23. I found it difficult that the RASTER manual was only available in English.	SD	D	N	A	SA
24. It is important that all RASTER documentation is available in Dutch.	SD	D ▲	N	A ▲▲▲	SA ▲

Accuracy and efficiency of risk identification

25. The RASTER method assists in quickly determining all large risks.	SD	D	N ▲▲	A ▲▲▲	SA
26. There exist large risks for unavailability of telecom services that were not found in this project.	SA	A	N ▲	D ▲▲	SD ▲▲
27. The RASTER method helps to ignore all small risks quickly.	SD	D ▲▲	N ▲▲▲	A	SA
28. The final risk list also contains risks that actually are not that important.	SA	A ▲	N ▲	D ▲▲▲	SD

Correctness of the outcome

29. I am confident that the results of the project are correct.	SD	D	N	A ▲▲▲▲	SA ▲
30. I can take responsibility for the results of the project towards my colleagues.	SD ▲	D	N	A ▲▲	SA ▲▲

Correctness of risk prioritisation

31. The RASTER method helps to determine risk priorities.	SD	D	N	A ▲▲▲▲▲	SA
32. The risks on the final risk list are shown in the right order (highest priority first).	SD	D	N	A ▲▲▲▲▲	SA

Effectiveness of the project

33. I am surprised by the results of this project.	SD	D ▲▲	N ▲▲▲	A	SA
34. Altogether, the project has made [organisation] (much less secure . . . much more secure).	MLS	Less S.	N ▲▲▲	More S. ▲▲	MMS

Figure 9.5 – (continued on next page)

Clarity of procedures

35. For me drawing and discussing the telecom diagrams was (very unnecessary ... very useful).	VU	Somewhat Unnry.	N	Somewhat Usefl.	VU ▲▲▲▲▲
36. I found the telecom diagrams (very unclear ... very clear).	VU	U	N	C ▲▲▲▲	VC ▲
37. For me estimating and discussing Frequency and Impact was (very unnecessary ... very useful).	VU	Somewhat Unnry.	N	Somewhat Usefl. ▲	VU ▲▲▲▲
38. I found the estimates of Frequency and Impact (very unclear ... very clear).	VU	U	N ▲	C ▲▲▲▲▲	VC

Technology Acceptance Model – Perceived Ease of Use

39. Learning to operate RASTER would be easy for me.	SD	D ▲	N ▲▲	A ▲▲	SA
40. I would find it easy to get RASTER to do what I want it to do.	SD	D ▲▲▲	N ▲▲	A	SA
41. My interaction with RASTER would be clear and understandable.	SD	D ▲	N ▲▲	A ▲▲	SA
42. I would find RASTER to be flexible to interact with.	SD	D	N ▲▲▲	A ▲▲	SA
43. It would be easy for me to become skilful at using RASTER.	SD	D ▲	N ▲	A ▲▲▲	SA
44. I would find RASTER easy to use.	SD	D ▲▲	N ▲▲	A ▲	SA

Technology Acceptance Model – Perceived Usefulness

45. Using RASTER in my job would enable me to accomplish tasks more quickly.	SD	D ▲▲	N ▲	A ▲▲	SA
46. Using RASTER would improve my job performance.	SD	D ▲	N ▲▲	A ▲▲	SA
47. Using RASTER in my job would increase my productivity.	SD	D ▲	N ▲▲	A ▲▲	SA
48. Using RASTER would enhance my effectiveness on the job.	SD	D	N ▲▲	A ▲▲	SA ▲
49. Using RASTER would make it easier to do my job.	SD	D ▲	N ▲	A ▲▲▲	SA
50. I would find RASTER useful in my job.	SD	D ▲	N ▲	A ▲▲▲	SA

Figure 9.5 – Questions and answers to the exit questionnaire for the second field test; each mark indicates one participant's answer. SD, D, N, A, SA stand for strongly disagree, disagree, neither agree nor disagree, agree, strongly agree respectively.

for Frequency and Impact were clear (questions 10–11), participants found it difficult to assign the most suitable estimate (questions 12–13); they were not sure whether the scales are really suitable, and the consensus arrived at often differed from their own estimate (questions 14–16). The time available for the project was somewhat tight, but sufficient (questions 17–20), and participants readily made themselves available (questions 21–22). Participants agreed that RASTER helps to discover large risks, but were less certain whether it was able

to weed out small risks (questions 26, 28); they doubted the efficiency of the process, especially for small risks (questions 25, 27). Still, they have confidence in the correctness of the outcome of the project (questions 29–32), but less so in its effectiveness (questions 33–34). Answers to question 30 show an outlier; we have reason to believe that this score was made in error. The process of creating diagrams and discussing vulnerabilities of components was found very useful (questions 35–38).

Group discussion

Afterwards, a one-hour discussion was held. The topics for this discussion were the same as for the first field test: 1) the RASTER methods and its strengths and weaknesses, 2) reliability of the outcome, 3) time and effort, and 4) determination of correctness. Below is a summary of the comments by the participants; unless indicated otherwise, the opinions below are the participants'. Because of time constraints, some of the issues could not be discussed in detail.

1) The participants found it difficult to create the diagrams, because diagrams show physical relations instead of logical relations, which they were more used to. They said that diagrams helped to gain insight, but tended to become complex. Diagrams allow for bundling of knowledge and expertise that would not have been possible in text-based representations. They suggested that it may be better to agree in advance on a certain depth and breadth of modelling. For example, the office network was modelled in detail, whereas provider networks were described at a high level; some IT infrastructure was modelled that extended beyond strictly telecommunication services. The more detailed the diagrams are, the more time has to be spent on their analysis. It would have been better, some participants thought, to have limited the project to a single high-level diagram, and a faster assessment without too many details.

The scales for Frequency and Impact were suitable. Having just three options forces one to make a choice; with more than three options the scores would probably cluster around the centre. Estimating the Impact was found to be easier than estimating the Frequency of vulnerabilities.

The participants were less confident about the correctness of common cause failure assessments, for both the clustering and the risk assessments.

2) Participants agreed that estimates are subjective, but saw no large issues with this, because the project team provides different views and group-discussions lead to consensus. The team composition lacked a representative from the waste water treatment department. Also, in some of the work sessions too few project members were present. Then again, the participants believed that the results of the project would not have differed.

3) The participants agreed that the time spent was sufficient; when more time had been available the results would not have been better. Had less time been available, then the analysis would have been less thorough. It is important to

have an experienced project leader to guide the team through the RASTER method and to monitor consistency. The participants remarked that it would have been better to shorten the lead time, and have multiple sessions per week. However, they also agreed that compressing the project into two or three consecutive full days would not have been useful: it is necessary to have time between sessions to research missing information and to “let matters sink in”.

4) Participants confirmed that they believed the project results to be correct. However, there have already been changes in the infrastructure since the start of the project, and some of the diagrams are already partly outdated. The main benefit of the project was that participants gained better arguments for their risk assessments, less so the risk identification. The final risk shortlist and the priorities of the risks were mostly as expected, and unsurprising to the participants. The fact that a group of mixed experts arrived at the same conclusions based on consensus makes that the recommendations carry more weight. It would be useful to repeat this project in a year’s time; lead time would be shorter than for the initial project.

9.3 Discussion and comparison

Since this is the second field test, its results can be discussed in contrast to the first field test. Also, the research questions for the second field test are answered.

9.3.1 The Raster method

Although the first team had ten telecom services and the second team only seven, the total number of components in the two projects was nearly identical: 205 in the first field test versus 204 in the second. Compare Figures 8.3 and 9.3. In the second project the diagram for maritime radios was trivial: in this service, ships communicate directly using independent radio sets; there is no central infrastructure. Excluding this service, the diagrams in the second project were clearly more complex than in the first. Whereas the first project shows on average 21 components per service (median 21), in the second project this was 36 (median 28). Furthermore, in the second project the services were in the process of being redesigned, and migrated from one telecommunication provider to another. The difficulty to discover the actual status quo further complicated the creation of diagrams.

The total number of vulnerabilities assessed was nearly identical: 832 single failures and 72 common cause failures in the first field test, versus 888 and 81 respectively in the second field test (Figures 8.4 and 9.4). This is to be expected, given that the number of components was nearly identical.

There was also overlap in the telecom services employed: both projects featured fixed telephony, mobile telephony, the national emergency telephony service, and private mobile radios. As the use of these services is different for

Figure 9.6 – Time spent (in hours) by the experts, excluding the experimenter, in each of the stages of the RASTER method. “Project general” includes the kick-off session, answering the questionnaire and the final discussion.

the two organisations, their diagrams in the two projects were not identical. However, the central architecture for these services was equivalent.

Project execution

The first field test required five work sessions; the second field test required nine. However, the total time spent on the projects is nearly identical (137 hours for the first project, versus 138 hours for the second, see Figure 9.6) and, as shown above, the size of the projects was also nearly identical. Seven experts participated in the first project versus five in the second project, meaning that the expert’s average contribution in the second project was higher (20 hours per expert on average in the first field test, versus 28 hours per expert in the second).

Despite these similarities there was a large difference in lead times: 7 weeks for the first field test, but 19 weeks for the second. When the three week break over the summer holidays and other idle time are excluded, the second project still took 3 weeks longer. See Figure 9.7 for an overview of the duration of the various stages. The difference in lead times does not originate from Stage 1 (initiation and preparation) or Stage 4 (evaluation); in both cases, these required one and two work sessions (and weeks) respectively. It is the risk assessment stages where a large difference is visible. In the first project the base diagrams were created in one work session, with minor revisions in the subsequent sessions. In total, Stage 2 (single failures) took three weeks. In the second project it took four work sessions before the diagrams were stable; overall, Stage 2 took five weeks. Stage 3 (common cause failures) took one work session and one week in the first project, and twice that much in the second project. There are three explanations for these differences.

First, the work load in the first project was shared between the core group and the plenary work sessions. The core group was able to prepare much of the output, for presentation to the entire group during the next work session. This approach was not possible in the second project, for practical reasons. As a result, most of the work had to be done during (plenary) work sessions. In Stage 2, in drawing the diagrams and assessing single failure risks, there are

Figure 9.7 – Duration (in weeks) of each of the stages of the RASTER method.

two kinds of activities: compiling the facts about physical components and their connections, and reflecting upon the diagrams and the risk assessments. It is more efficient to perform the first activity in a dedicated in-depth expert session, rather than during plenary sessions.

Secondly, it was far more challenging in the second project to create the telecom service diagrams, because the services were more complex. There were more locations, and more telecommunication technologies. For example, the telemetry service (Figure 9.2) makes use of DSL, fiber-optic, ISDN, analog telephony, 2G modems, 2G and 3G mobile data, VPN, SMS, and paging all for a single telecom service. On top of this, there was recurrent confusion about the actual situation, because services were being migrated to a new provider.

Thirdly, the second project team found it more difficult create diagrams. In the group discussion they noted that telecom service diagrams in RASTER, by design, show physical components and physical connections, whereas they are more used to thinking about telecom services as abstract logical components and information flows. The RASTER diagrams forced the experts on information and communication technology to think about their infrastructure in a way that they were less used to. In the group discussion one participant raised the suggestion to limit the project to a single high-level diagram. Such a simplification may work in projects where services are tightly coupled, and where most risks are low with few high-risk exceptions. If services are tightly coupled then there will be many components and functions that are shared between services. As a result it will be more difficult to separate the diagrams for these services, and a single diagram may be more efficient. However, such a single diagram tends to have many components and can therefore become complex. Complexity can be reduced by keeping details 'hidden' within Unknown links (the cloud shapes in diagrams). The disadvantage of this is that availability risks within those Unknown links are more difficult to identify. If most risks are low and high-risk exceptions can be located with reasonable accuracy, then the disadvantage of using many Unknown links is limited. However, many organisations use services that are loosely coupled, with risks cannot be located easily. Use of a single large diagram will then not be helpful for accurate risk assessment. In the project at Safety Region Groningen, for example, such an approach would not

have been effective; the risks that were identified in that project would likely not have been found when only a single high-level diagram had been drawn.

In the group discussion it was said that it would have been useful to have an additional expert on waste water treatment in the project team. Team composition was addressed both at the start of the project, and later during project progress review. At those moments, the team agreed that no further participants were required, as the necessary knowledge was already present in the team. The participants believed that the project outcome would not have been different, if a waste water treatment expert had been present. It therefore appears that the participants collectively were able to supply the necessary expertise, and that an additional expert was not required.

In the group discussion it was also remarked that the common cause failure assessment may have been inaccurate and incorrect. Stage 3 (common cause failures) was carried out just before the summer break, with only a few participants present. Most project members were unable to give this stage their full attention. This will have prevented them to contribute and carefully review the assessments. However, such circumstances are a matter of project management, and are not indicative of an issue with the RASTER method.

Project results

In addition to the higher lead time and complexity of the second project there is another notable difference. The second team's Impact assessments were less severe, and the number of high Frequency and high Impact assessments was lower than in the first field test (see *Figure 9.8*). The reason for this, according to the participants, is that the operations of WHA have been designed with redundancy and resilience in mind. For example, telemetry provides much of the essential information for water level management, but disruptions of a few hours can be tolerated without danger. When remote sensors come back on-line, they will transmit all queued measurements. If disruptions are to be measured in days, then field personnel will record water levels manually, and relay this information to head office via private mobile radio or mobile phone. Even sea barriers can be closed manually in emergencies.

This lower overall risk level could also be observed during Stage 4 (evaluation). Whereas the longlist in the first field test contained 31 High risks, the second field test only contained 11.

Questionnaire outcome

The scores given by the two teams are often nearly identical (see Appendix B.3). For example, there is very little difference in their answers on ability to contribute (question 5–7), attitude to consensus (question 8–9), clarity of the scales (question 10–11) and time required (question 17–21).

A notable exception is the set of answers on applicability of the scales (questions 12 to 16). The teams from the second field test scored more negative,

Meaning of letter codes: U stands for extremely low (Unlikely or Ultra low), L for Low, M for Moderate, H for High, V for extremely (Very) high, X for unknown, A for Ambiguous, and - for undecided.

Figure 9.8 – Distribution of Frequency and Impact assessments for single failures the in the first (light grey) and second (dark grey) field tests.

especially on questions 12 and 13 (hesitating between adjacent Frequency and Impact scores). Apparently, the second team had more difficulty in coming to a consensual decision when assessing Frequency and Impact, compared to the first team. The reason for their hesitation was not that the scales themselves were unclear, as indicated by the near-unanimous scores to questions 10 and 11. Where the first team was quite confident about the applicability of the scales (questions 14 and 15), the second team was somewhat less certain. A possible explanation for the difference between the two teams may be that the second team's Impact assessments were much less severe. It is possible that sifting through a seeming endless list of low risks gave the second team a sense of lack of direction. Unfortunately, it is only *after* completing the risk assessments that it becomes clear that most risks are small.

This conjecture may then also explain the results of another pair of differences. Question 26 asks whether there exist large risks that were not identified during the project. The second team, with hardly any High risks and overall lower assessments, was more confident that this was *not* the case (a more positive outcome). However, that same team was more negative on the question whether small risks were quickly ignored; they had to assess many more Low-impact risks.

9.3.2 Research questions

Recall that there are three research questions:

- Q1 Does RASTER produce, in practice, risk evaluations that are correct? That is, are all relevant risks included (completeness), are low risks excluded (conciseness), and are risks presented with the right priorities?
- Q2 What are the costs and effort involved in execution of RASTER?
- Q3 Are the target users willing to use the RASTER method?

Q1: Correctness

The participants' answers to the questionnaire (questions 29–32) show that they have confidence in the correctness of the project results. As for the aspect of completeness, the participants were certain that no large risks have been overlooked, but less certain about the efficiency of this activity. They were divided on the conciseness of the outcome: some participants indicated that the final risk list might contain too many unimportant risks. The activity of eliminating small risks was perceived to be slow. Despite these reservations, the participants were positive about having found the right risk priorities. Overall, it can be concluded that the participants believe the results to be correct, but somewhat inefficient, especially for small risks.

Q2: Required effort

The five participants spent on average 28 hours on the project (see *Figure 9.6*). The questionnaire results and group discussion both indicate that the time for the project was sufficient. The participants were aware that the project required a significant portion of their time, but were able to make themselves available for the project without too much difficulty. Overall, it can be concluded that the time required is acceptable.

Q3: Disposition to use

As in the first field test, we used the original Technology Acceptance Model to assess the participants' disposition to using the method in other projects. Answers to the questions on perceived ease of use were inconclusive (*Figure 9.9*). This topic was also remarked upon in the group discussion; participants said that an experienced project leader is essential and that they would not have been able to employ the RASTER method without help. This opinion will have been influenced by the large lead time of the project, as discussed in the previous subsection. It was not uncommon for participants not to work on the project for three weeks between work sessions. A more compact project planning may have resulted in more positive answers in the ease-of-use section.

The answers for perceived usefulness (*Figure 9.10*) were slightly more positive but still somewhat inconclusive, and markedly different from the scores from the first field test. These opinions are probably influenced by the large number of Low risks, and by the fact that there were few High risks, as discussed in the previous subsection.

9.4 Peer review

We asked the project team members to participate in a limited peer review. The goal of the review was twofold:

- to validate reliability of the RASTER method, and
- to validate the correctness of its outcomes.

Figure 9.9 – Scores on the TAM questions on Perceived Ease of Use.

Figure 9.10 – Scores on the TAM questions on Perceived Usefulness.

In Chapter 7 a method is called reliable if it can be repeated with the same results. We proposed a method for experimental validation of reliability, and indicated that for practical reasons it cannot be applied using professionals. Therefore, we used peer review instead. Participants from the second field test were asked to review the products of the first project. In the context of peer review, ‘correctness’ means that results are based on accurate input data, consistent with existing knowledge, and derived according to the RASTER procedures.

A peer review has two main disadvantages. First, a full review takes a lot of time; given the already long lead time of the project we were only able to dedicate a part of the final project meeting to the review. Secondly, a full review requires domain knowledge about the target organisation. The first project team included specialists from emergency medical care, police, civil support, and fire services; the second project team did not possess this expertise. This restricted the possible scope of the review.

To minimise our capacity claim on the participants and to limit our questions on areas of shared expertise, we carefully selected partial results from the first project. This selection excluded the results of Stage 1 (identifying services, actors,

Figure 9.11 – Two equivalent representations of component instances with alternative connections. The representation on the right was found to be more clear.

etc.), the drawing of telecom service diagrams, treatment recommendations, and any other area that requires domain knowledge about Safety Region Groningen that was not available in the second project team. We selected the following four topics: understanding the diagrams, correctness of single failure assessments, clustering and assessments for common cause failures, and assessment of social risk factors. For each topic we selected parts of the first project's results (e.g. a diagram, two or three single failure assessments) and asked specific questions about these parts. The results are described and discussed below.

Understanding the diagrams. We wanted to know whether participants could explain key elements of a telecom service diagram, such as the path of communication between two actors, and the intended meaning of multiple paths. Participants were able to explain the flow of communication and the purpose of components, and correctly identified built-in redundancy in parallel connections.

When a diagram component represents not a single instance but a collection of instances (for example 'all fire stations') then the meaning of multiple connections to this component can have two meanings. Either it means that every instance has all connections, or that each instance has one of these connections (see Figure 9.11). Participants were of the opinion that only the first meaning is correct; if the second meaning is intended, an alternative diagram should be used. This disambiguation is not more than a diagram technique; the risk assessments will be identical in both cases.

Recall that several telecom services occurred in both the first and the second field test. This allowed us to compare the two variant diagrams developed by the teams. We selected mobile telephony for this comparison. The participants found no substantial differences between the two diagrams. The WHA-version showed somewhat more detail, e.g. to explicitly show the wireless connection between smart phones and the mobile network, but participants disagreed on whether this added useful information to the analysis or not.

The participants suggested that, in order to make diagrams more understandable, additional information could be recorded. For example, a small note could

indicate the starting point of communication. It would also help understanding if a concise explanation of key points was included in the diagrams. Both suggestions are already supported by the current version of the RASTER tool; it is possible to insert coloured notes containing arbitrary text.

Single failure assessments. Another service appearing in both projects is private mobile radio. The participants were hesitant to review the assessments made by the first project team, indicating that they lacked the required domain knowledge, especially for impact assessments. Respect for the professional integrity of the first project's team members may play a role here. When pressed, they were able to compare the assessments against their own situation, noting similarities and differences in their organisations that would explain the difference in assessments. They also pointed out risk assessments that, in their opinion, were unjustifiably high. In addition to domain knowledge, participants also indicated subjectivity as a factor in differences between assessments. They reiterated the importance of a balanced group composition in order to moderate the effects of subjectivity.

Common cause failures. We selected the general vulnerabilities of malfunction and configuration errors by end-users and maintenance staff. The participants found it difficult to comment on the clustering criteria and on the cluster composition. As explained in the previous section, analysis of common cause failures did not receive sufficient attention in their project. This will also explain their incapacity to critically discuss the common cause failure assessments of the first project.

In addition to the two vulnerabilities above we also selected flooding. Flooding is, of course, of special interest to the staff of the water board. Here participants were able to make useful suggestions on alternative clustering properties (height above sea level instead of physical proximity), and on additional clusters. For example, additional clusters could have been created for flooding of river banks, or flooding after heavy rainfall.

Evaluation. We asked participants to reflect on the assessment of social risk factors, as performed by the first team. Participants saw no omissions nor unnecessary items. They were struck by the emphasis on public accountability and danger of media attention, issues that were less important in their own project.

When asked, the participants were unable to name large risks that were not mentioned and addressed in the first project. Again, they stressed the need to have representatives from all disciplines on the project team, to prevent such omissions.

Participants observed that both organisations are dependent on external partners and suppliers. They also speculated that the RASTER method could be adapted to a wider range of risk assessments. For example, they suggested that

RASTER or a modified method could be applied to business processes in general. Also, cyber security risks could be explored using the RASTER method and tool. The project team would in this case have to be extended with a cyber security expert.

Conclusion

Using a limited peer review we gathered additional data on the reliability and correctness of the RASTER method. The participants were able to correctly interpret selected diagrams from the first project, even when they lacked domain knowledge.

As for reliability, the second project team did not indicate steps in the RASTER method where their approach would have obtained different results. Diagrams for services that appeared in both project were similar, with differences based on the particular needs of both organisations. Participants indicated a preference for a particular diagram technique, but this difference would only improve the understandability, and not lead to different risk assessments. In the few examples presented during the review, single failures and common cause failures would not have been assessed differently. However, when presented with an area where the participants' domain knowledge exceeded the knowledge of the first project team, several enhancements and additions were suggested. These findings support the statements in previous chapters about generally low reliability of expert assessments, and the need for good justifications for assessments and risk treatment decisions.

Correctness means that results are based on accurate input data, consistent with existing knowledge, and derived according to the RASTER procedures. The participants had questions on some of the assessments of the first project team, but did not find reasons to label any of their results as incorrect. The review does not provide arguments to change previous statements on the correctness of RASTER.

9.5 Lessons learned

The previous three chapters end with a section on design improvements. This section does not. The findings in the field test at the Water Board Hunze en Aa's teach us a lot about managing a RASTER project, but do not lead to improvements in the method or tool themselves. There are, however, a number of lessons to learn.

Project lead time. The project results would have been more accurate and perhaps more complete if the lead time had been shorter. The total time spent was sufficient, but spread out over too many weeks. Although it was noted during project review meetings that the planned deadlines had to be extended, the author underestimated the negative impact thereof. Participants did not

gain the exposure necessary to become fluent in the RASTER method. In future, it is necessary to stay close to a 5 to 6 week lead time. Also, it is necessary to obtain commitment from the project sponsor on the availability of project members, to ensure that all project members are present during plenary meetings.

Organising work sessions. The first project made use of a core team as part of the full project team. The core team could collect the facts about telecom service components and their composition. This is an activity in which, in general, the full project team cannot make a useful contribution. An in-depth experts-only session will be more effective and efficient in creating the initial diagrams. The full project team is necessary to reflect upon the diagrams and risk assessments. Based on these reflections the assessments can then be completed in a smaller expert team, possibly a different team than the one creating the initial diagrams. This two-tier organisation of work sessions would have made better use of the scarce time of participants, and will have a beneficial effect on lead time.

Small risk projects. The second project was carried out in an organisation characterised by high redundancy, high resilience, and low availability risks. The RASTER method will perform equally well for such organisations as for organisations exposed to higher risks. However, it is important for the project leader to realise that project members may become somewhat annoyed or frustrated with perceived lack of progress. In such situations, it will be better to spend less time on single failure analysis, and more on common cause failures. For example, instead of assessing every single vulnerability the team may decide to assess only those vulnerabilities that may result in a High risk score, and leave all others unassessed, thus saving time and effort. As it may not be apparent at the start of the project for this to be the case, the project leader should be aware of this possibility, and prepared to switch to a different approach if necessary.

9.6 Conclusion

Some general conclusions can be drawn from the results of both field tests.

Assessment of risks with high uncertainties requires a large amount of expert knowledge. Facing such risks, decision makers depend on a team of risk assessment and domain experts. Different assessment teams will arrive at assessments that are similar but not identical. If the assessments are both correct, as can be expected from well-motivated experts, the differences in assessments will be supported by the difference in arguments for the assessments as provided by both teams. In either case, decision makers can use the risk assessment as their justification for making risk treatment choices that are both scientifically well-grounded and supported by most stakeholders. In general it will be impossible to state whether one team's assessment was incorrect, or the other team's assessment was more correct. Given limited resources, decision makers

are confined to the best possible assessment and justifications available at the time.

Given the need for an assessment team composed of experts from various backgrounds, information sharing and discourse within the team will be difficult but important. Rich graphical models can facilitate shared understanding and discussion about architectures, not only within one team but also between teams. This shared understanding is limited however by the need for specific domain knowledge. In general, peer review of risk assessments is therefore restricted to aspects of the assessment that are general or do not require domain expertise.

Conclusions

This chapter summarises the findings leading up to this conclusion, answers the research questions, and provides an outlook to further research.

10.1 Summary

As stated in the Introduction, this research has an overarching practical aim: not so much to further theory or the state of the art, but to solve a practical problem that actually exists in current practice.

We developed a new risk assessment method for risks to availability of telecommunication services as used by crisis organisations. We stated our requirements based on experience and current practice. Our motivation for developing a new risk assessment method is that current practice and methods do not fully satisfy these requirements. We designed an initial version of our method, and argued that this design satisfies most of our requirements. To show that the method satisfies all requirements, and to validate that the method can be performed in practice, we conducted a case study. Based on the lessons from the case study we improved the method; this cycle was repeated a few times. Finally, we refined the design such that the method could be performed from beginning to end in what we believed to be reasonable time.

To validate the reliability of the method, we conducted an experiment in which six groups performed the core analysis part of the method independently and in parallel. Comparing the results, we noticed a relatively low agreement in assessments. In part we could attribute those differences to artifacts of the experiment, but one cause was internal to the method. We improved the method in that respect. Still, reliability of expert risk assessments is low, and experts retain a large responsibility for their results. Our later field tests show that experts are aware of subjectivity, and are willing to take this responsibility.

Finally we tested the method in the field, as technical action research. In these tests, we assisted an actual crisis organisation in assessing their unavailability

risks of telecommunication services. We conducted two such field tests. In both tests the method could be applied by local experts to real telecom services within reasonable time. Furthermore, the practitioners endorsed the method: they expressed their agreement with the outcomes of the projects and affirmed its usefulness. The first field test led to improvements to the RASTER tool; neither field test led to changes to the RASTER method itself. In the second field test we asked the experts to perform a limited peer review of the results of the first project. The review did not demonstrate a need for improvements to the RASTER method. The results of both field tests were presented to management, and most of the treatment recommendations were actually put into practice.

10.2 Review of research questions

The Introduction presented the three main research questions for this thesis. The subsequent chapters elaborated our answers to these questions. These answers are summarised below.

Research question 1

The first research question asked for a justification for this research.

RQ1: Are current risk assessment methods able to adequately assess availability risks in our problem context?

RQ1-a: What existing risk assessment methods can be applied to telecommunication services and their use in crisis organisations, and what are their properties?

RQ1-b: What risk assessment methods are currently used by crisis organisations?

RQ1-c: What are the requirements for risk assessment methods in our problem context?

RQ1-d: What risk assessment methods do match those requirements?

RQ1-a was answered in Chapter 3. We looked at general principles of risk assessment methods in industrial settings, at technology assessment, and at methods for assessing information security in ICT systems. Then, we investigated a selection of specific methods from each of these settings, as examples illustrating the properties of the methods in these settings.

RQ1-b was answered in the same chapter. Most organisations, if they perform a risk assessment at all, use an informal ad hoc process. One notable formal method is being used in Dutch practice: a continuity plan framework developed for the Safety Regions. This method leads to useful results, but has analytical shortcomings that weaken its risk assessments.

RQ1-c was answered in Chapter 4. We found that requirements on risk assessment methods are related to the three different scopes in which the risk assessment is performed. As the scope is expanded from the target of assessment, to the decision maker, and to the stakeholders, additional risk factors become relevant. When the scope is restricted to the target of assessment, risk factors are mostly objective, quantifiable, and easy to influence with design choices. When the scope is expanded additional factors are often are subjective, qualitative, and not easily influenced. In addition to these sets of risk factors, risk assessment methods in our problem context should match the specific properties of the problem domain. Information on telecommunication services will be complex and incomplete. Busy experts from very different educational and professional backgrounds need to collaborate. Most importantly, methods must effectively assist decision makers to choose risk treatment options, and to explain those choices to all stakeholders.

Finally, Chapter 4 answered research question RQ1-d. None of the investigated existing methods satisfies all requirements. Furthermore, because of their properties it will be difficult to design modifications to them to make them suitable to our domain.

We can therefore answer question RQ1 negatively: methods used in practice and available methods in general have shortcomings and do not sufficiently fulfil the requirements of our domain. Development of a new risk assessment method is therefore warranted.

Research question 2

In RQ2 we address the design of a new risk assessment method to fit the requirements from RQ1-c.

RQ2: How can we design a risk assessment method for availability risks in our problem context?

Although RQ2 is stated as a question, it is actually a design problem. We applied principles and theories from design science to iteratively specify a treatment (a version of our new risk assessment method), validate that treatment in an experiment, and improve the treatment based on the outcome of the experiment. We repeated validation and improvement until we were satisfied that all requirements were fulfilled. These repeated design steps were described in Chapters 5 to 9.

Important general design recommendations were for tool support and rich graphical architecture models, and for explicit feedback mechanisms and review steps to ensure consistency. These design recommendations help to ensure that risk treatment decisions are founded on the best possible arguments, given limited resources.

Research question 3

In RQ3 we validate the new method against reliability, correctness, and the specific requirements derived from question RQ1-c.

RQ3: What is the contribution of our new risk assessment method?

RQ3-a Is the new risk assessment method feasible: can the method be performed in practice?

RQ3-b Is the new risk assessment method reliable: can the method be repeated with comparable results?

RQ3-c Is the new risk assessment method an improvement over current methods?

RQ3-a was addressed in Chapter 6. The initial design proved not to be feasible in lab experiments. We improved the steps of the method, and most notable introduced the RASTER tool. This iteration ended when we concluded that in a lab environment the method could be executed from beginning to end within reasonable time. A field test of feasibility had to wait until other properties of the method were tested in a lab environment, as we did not want to impose on experts until we were reasonably certain that our method would work. Practical feasibility was tested in two field tests, as described in Chapters 8 and 9. Both field test confirmed that, although time consuming, the RASTER method is not overly imposing on experts' time and sufficiently useful to warrant its expenditure.

RQ3-b was answered in Chapter 7. We found that RASTER was not quite repeatable in a lab environment. These results are in line with previous research showing generally low inter-rater agreement in methods that are dependent on expert judgements. However, we argued that in a field environment repeatability would likely be better, and later tested this in a peer review. Furthermore, we were able to explain differences in outcomes partially by weaknesses in the method. These weaknesses were treated in a subsequent redesign of the scales used for experts' estimates. As a result, the RASTER is as reliable as can be expected from a method that draws heavily on the expertise and experience of practitioners. We had experts conduct a peer review to test reliability in a field environment. Peer review proved difficult, as risk assessments in practice are heavily dependent on specific domain knowledge. This review did not reveal any shortcomings, nor suggestions for improvement.

RQ3-c asks whether RASTER improves upon existing methods. Although RASTER borrows many ideas from existing risk assessment methods (e.g. the use of checklists, combination of qualitative likelihood and impact assessments), it has some distinguishing features.

- RASTER has been designed specifically to assess telecom service availability

risks. Its modelling language (telecom service diagrams) directly represent physical components of which telecom services are composed.

- RASTER has been designed for crisis organisations: public organisations that must not only manage risks, but must also account for risk treatment decisions, and take all stakeholders' preferences into account.
- RASTER is a practical method with step by step instructions that does not require customisation before it can be applied. It is documented in an application manual, and is supported by a tool that has emerged from several rounds of practical testing.
- The risk treatment recommendations created with RASTER are based on traceable arguments. Each recommendation can be traced back, down to individual Frequency and Impact assessments if necessary.
- RASTER handles missing information and complexity. It does not require a complete model of the entire target. Assessments can be left open or be explicitly marked as Unknown, and unknown links can be used to hide areas that are not of interest, or of which the internal composition is unknown.

To our knowledge this combination of features is not available in other risk assessment methods.

The most important feature of RASTER, and our the most important answer to RQ3, is that outside the lab, in a professional environment, RASTER has been used by two organisations to actually find and implement risk reduction measures. RASTER is more than an academic research topic, but can make a difference in practical circumstances.

10.3 Generalisation and limitations

The research question in this thesis is targeted at crisis organisations; case studies and field tests have therefore been performed at crisis organisations. In this section we speculate on whether RASTER could be of benefit to other types of organisations or risks, and discuss some possible limitations.

The requirements developed in Chapter 4 are equally applicable to other public end-user organisations, with the exception of requirement R6 *Exceptional circumstances*. However, actually experiencing disasters and other exceptional circumstances is not mandatory. Therefore, nothing prevents the use of RASTER in organisations that do not operate in exceptional circumstances. Use of RASTER can very likely be generalised to any organisation with a public responsibility and obligation to account for risk mitigation measures.

If an organisation has no such responsibility and obligation, the RASTER method could still be useful. Social risk factors would, in this case, likely not contribute to the risk prioritisation nor to risk treatment recommendations. In

this case it would probably be better to skip and omit the assessment of social risk factors in Stage 4 (evaluation).

RASTER may be of limited use in organisations where all staff (end-users and maintenance personnel) are highly homogeneous. Stage 2 (single failures) and Stage 3 (common cause failures) use a discursive, consensus-based assessment of vulnerabilities to reconcile different views from different areas of expertise. This may be inefficient when it is not needed to arrive at a common understanding. Also, perhaps more specific and precise scales for Frequency and Impact can then be used; the analysts may be able to come to a more detailed consensus. When analysts agree closely on assessments, it may even be better to replace the qualitative scales with quantitative ones. This would mean a fundamental change to the RASTER method.

RASTER will also be of limited use to organisations without end-users, such as telecommunication operators. Analysts in this sector would have no need for the five high-level component types used in RASTER (actors, wired links, wireless links, equipment, and unknown links). In their operations, a much more detailed modelling of infrastructures will be useful, also because the analysts will all be telecom experts with very similar and high-level expertise.

This research was carried out in the Netherlands, and validated at local organisations. RASTER does not depend on Dutch crisis structures, and should be applicable to other countries as well.

RASTER assesses telecom service availability risks. It is a typical property of telecom services that they are composed of services offered by several independent, competing companies. If this is not the case, e.g. when all telecom services of an organisation are offered entirely by an internal ICT department, RASTER may be inefficient. When knowledge about the underlying telecom infrastructure is complete and available to the team, other methods of risk identification may be available that require less effort.

The five component types are specific to telecommunication. Other kinds of technical infrastructures use different component types. Perhaps RASTER can be extended to these infrastructures, using the component types that are specific to that infrastructure. However, this is most likely only useful when these infrastructures are offered, like telecommunication, by multiple independent and competing companies. For example, we think it unlikely that RASTER could be adapted to the Dutch railroad network, for which a single company is responsible. On the other hand, perhaps RASTER can be adapted for electrical power distribution networks, since these are moving towards locally generated energy from renewable sources (wind turbines, solar-panel covered roofs).

RASTER has been developed to assess availability risks, without limit to specific kinds of hazard or threat. Unavailability due to cyber incidents is therefore within the scope of the method. However, RASTER does not closely analyse deliberate attacks; it does not assess aspects such as the value of

information or ease by which a weakness in defences can be exploited. RASTER's telecom service diagrams do not model functional and behavioural relations between components, which is necessary for a thorough security risk assessment.

Further research is necessary to test whether RASTER can be used or adapted for the above targets.

Applicability

Handling change is an important aspect of risk management. Risk targets are constantly extended and modernised, especially so in telecommunication and information technology. Risk assessment should therefore be a continuous process. In both field tests the sponsors wanted to assess changes in risk to their services. RASTER does not model modifications to telecom services. However, when a risk assessment is repeated on a risk target that has since been upgraded, the second project can reuse many of the results of the first project, and the amount of work for the second project will be much less than for the first. Analysts will then compare the two shortlists, to determine the effect that changes have had on availability risks. Proposed changes can be assessed beforehand and compared to the existing situation in the same way.

RASTER is a bottom-up method, where risks to a telecom service are derived from risks to its parts. Other methods, such as the Continuity plan framework discussed in Section 3.2.2, are top-down: they investigate how critical business processes are dependent on telecom services, and how reliable these services are. There is merit in both approaches. Top-down methods can be aligned more closely to organisational priorities, but may overlook technical details about reliability of components. Bottom-up methods tend to be more detailed and more accurate on technical aspects, but may be less efficient when time is spent on telecom services that are less relevant to the organisation. Top-down and bottom-up methods can be combined; the results of an initial top-down method can be used to determine the scope of a subsequent bottom-up assessment of the most critical telecom services. It is up to the sponsor to decide what kind of risk assessment will provide the information that is most relevant to the organisation's current questions. Analysts performing a RASTER analysis should be aware that a detailed analysis of less important telecom services may not be the best use of their time.

In general, telecom services are dependent on infrastructures that are provided by third parties, and that are entirely out of the control of the organisation. In both field tests the organisations were dependent on commercial telecom operators for several of their services. This dependency is a risk that typically can only be avoided at great cost; there are clear financial benefits to buying services from large, specialised providers. Furthermore, relying on alternative infrastructures under your own control introduces its own risks, that may well be larger unless your organisation is capable of maintaining those infrastruc-

tures at the same level of quality and skill as commercial operators can. In practice, many organisations have no real alternative to relying on other telecom infrastructures. They can, however, reduce availability risks to components under their own control. RASTER has been shown to be a useful tool to that effect.

10.4 Outlook and further research

Crisis organisations can now use RASTER to create risk treatment recommendations that make their organisations more resilient, and better prepared to handle availability incidents. One limitation of the research presented in this thesis is that both its field tests have been performed by the author. It would be very useful to perform field tests with an other suitably trained person as the project leader. This may uncover undocumented assumptions or other deficiencies in the RASTER application manual, or even disclose limitations of the method. At Radiocommunications Agency Netherlands we aim to hold ‘train the trainer’ sessions, to transfer knowledge and experience to a wider pool of potential project leaders.

Of particular interest is the question how much telecommunication knowledge the project leader must have. The analysts will come from a wide range of backgrounds: some will have extensive technical expertise, some will have broad domain knowledge instead. Ideally, the project leader should have experience with applying the RASTER method, but no prerequisite knowledge of telecommunication or the end-user domain. Only further field tests with other project leaders can test the hypothesis that experience with the RASTER method is the only requirement.

The development of course material is also a necessity if the RASTER method is to be put to wider use. For use outside the Netherlands, it will also be necessary to provide translated versions of the RASTER application manual and tool. The RASTER tool can probably be localised with minor effort; the translation from English to Dutch has already laid the necessary groundwork for this.

10.5 Final remarks

The mission of Radiocommunications Agency Netherlands is to assure the availability of modern and dependable telecommunication in and for the Netherlands. To achieve this mission, the Agency works on both improving the reliability of telecommunication services and on increasing the resilience of their users. These two activities complement and reinforce each other. The programme *Telekwetsbaarheid* (“tele-vulnerability”) encompasses a number of research and educational activities to further end-user resilience. The RASTER method will be an important tool within this programme, and we expect that RASTER will be actively used and developed further.

Appendices

Krippendorff's alpha

This appendix is based on the author's previously published technical report "Testing reliability of Raster – report of experiment with Kerckhoffs students", TR-CTIT-14-05.

This Appendix describes the calculation of Krippendorff's alpha for ordinal data; for a detailed description see Krippendorff [99]. In particular, it shows and resolves an issue with alpha that arises when comparing alphas of two subsets of the data.

A.1 Computing alpha

Krippendorff's alpha can be computed for any data in which r raters assess N units. For example, in the reliability experiment described in Chapter 7 there are 6 groups of students assessing 276 units of likelihood and impact assessments. In this case each group is a single rater. Raters normally assess every unit, but when raters can abstain then the total number of scores can be less than $r \times N$. The discussion below is restricted to assessments on an ordinal scale; with other kinds of scale the calculations would be similar.

Krippendorff's alpha is defined as:

$$\alpha = 1 - \frac{D_o}{D_e} \quad (\text{A.1})$$

where D_o (the *observed* amount of disagreement) is the average difference between scores within units, and D_e (the *expected* amount of disagreement) is the average difference between all scores in the data: the disagreement if raters scored randomly. If the raters have perfect agreement, then observed disagreement is 0 and alpha is 1. If the raters' scores are indistinguishable from random scores then $D_o = D_e$ and alpha is 0. If $\alpha < 0$, then disagreement is larger than random disagreement.

The disagreement can be calculated from the averages as described above, but is more often based on the observed and expected coincidence matrices of

the scores. Elements in this matrix indicate the relative frequency in which a pair of scores occurs within units, or across all units, respectively. With U the set of all units, o_{ck} the observed coincidence of the pair of scores c and k , and e_{ck} the expected coincidence of c and k ,

$$\begin{aligned} o_{ck} &= \sum_{u \in U} \frac{\text{number of } c\text{-}k \text{ pairs in unit } u}{\text{number of scores in unit } u - 1} \\ e_{ck} &= \frac{\text{number of } c\text{-}k \text{ pairs in all units } U}{\text{number of scores in all units } U - 1} \end{aligned} \quad (\text{A.2})$$

Element o_{ck} is proportional to the number of times that the two scores c and k were scored in the same unit; it is defined in such a way that $\sum_k o_{ck} = n_c$, where n_c is the number of times that score c appears in the data. Coincidence matrices are symmetrical around their diagonal, so $o_{ck} = o_{kc}$ and $e_{ck} = e_{kc}$.

Element e_{ck} is proportional to the number of times that the two scores c and k would appear if assessment were completely random; $\sum_k e_{ck}$ also adds up to n_c . Random assessment means that raters blindly assign a score, while observing the relative frequency of scores (the number of times that each score actually appeared in the data). With n the total number of scores ($n = \sum_c n_c$), e_{ck} can be written as

$$e_{ck} = \begin{cases} \frac{n_c(n_c - 1)}{n - 1} & \text{if } c = k \\ \frac{n_c n_k}{n - 1} & \text{if } c \neq k \end{cases} \quad (\text{A.3})$$

In the calculation of D_o and D_e pairs are weighted to account for the ordinal distance between their scores. A pair of two adjacent scores would carry a lower weight (and hence contribute less to the amount of disagreement) than a pair of extremes from both end of the ordinal scale. Function δ_{ck}^2 is the *difference* between scores c and k . If $c = k$ then $\delta_{ck}^2 = 0$. For ordinal data, when $c \neq k$:

$$\delta_{ck}^2 = \left(\sum_{g=c}^k n_g - \frac{n_c + n_k}{2} \right)^2 = \left(\frac{n_c}{2} + \sum_{g=c+1}^{k-1} n_g + \frac{n_k}{2} \right)^2 \quad (\text{A.4})$$

Using the coincidence matrices, D_o and D_e can be written as:

$$D_o = \frac{1}{n} \sum_{c,k} o_{ck} \delta_{ck}^2$$

$$D_e = \frac{1}{n} \sum_{c,k} e_{ck} \delta_{ck}^2$$

Combining all this into equation (A.1) gives:

$$\alpha = 1 - \frac{D_o}{D_e} = 1 - \frac{\sum_{c,k} o_{ck} \delta_{ck}^2}{\sum_{c,k} e_{ck} \delta_{ck}^2} \quad (\text{A.5})$$

It is important to note that D_o , D_e and δ_{ck}^2 are all defined using the relative frequencies of scores as they appear in U .

A.2 Alpha over subsets

Suppose that there are two subsets of U (say, U^l and U^d) and that we want to verify whether the reliability of subset U^l differs from that of U^d . For example, U^l may be all assessments related to laptops and U^d all assessments related to desktops, and we would like to know whether the raters agreed more on one type of component. We would compute α^l over U^l and α^d over U^d and we would *like* that

$$\alpha^l < \alpha^d \stackrel{?}{\iff} \text{inter-rater reliability in } U^l < \text{inter-rater reliability in } U^d \quad (\text{A.6})$$

The problem here is that α^l is calculated using the relative frequencies of scores appearing in U^l , whereas α^d is calculating using different relative frequencies. For external reasons (e.g. the properties of laptops), the scores appearing in U^l may be rare in comparison to the entirety of U , whereas the relative frequencies in U^d may be comparable to those in U . α^l would be computed relative to the ‘distorted’ relative frequencies from U^l while U^d would be computed using regular frequencies. α^l might be ‘inflated’ or ‘depressed’ as a result, and our wish (A.6) above may or may not hold [98].

A solution is to compute expected disagreement and the weights δ_{ck}^2 on the basis of relative frequencies observed in the entire dataset U , not relative to their own, possibly biased, data-subsets:

$$\alpha^l = 1 - \frac{D_o^l}{D_e} = 1 - \frac{\frac{1}{n^l} \sum_{c,k} o_{ck}^l \delta_{ck}^2}{\frac{1}{n} \sum_{c,k} e_{ck} \delta_{ck}^2} = 1 - \frac{\frac{n}{n^l} \sum_{c,k} o_{ck}^l \delta_{ck}^2}{\sum_{c,k} e_{ck} \delta_{ck}^2} \quad (\text{A.7})$$

Compare this to equation (A.5). The δ_{ck}^2 and e_{ck} values as calculated for the overall α over U can be reused when calculating alphas of subsets, but the amount of observed disagreement must be scaled by a factor n/n^l .

A.3 Example

In this example there are three raters called A, B, and C; there are 14 units, assessed by the raters on an ordinal scale containing 5 levels $\langle U, L, M, H, V \rangle$. The results of each of the rater’s assessments are:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
A	L	H	M	L	M	M	U	V	U	L	M	M	H	H
B	M	H	L	U	M	M	L	H	L	L	L	L	M	*
C	M	H	H	L	M	H	U	H	H	L	M	M	*	*

In this table, the symbol * stands for missing data. For example, the odd-numbered scores could be Frequency assessments, and the even-numbered scores could be Impact assessments. The three raters then assessed likelihood and impact on seven RASTER vulnerabilities. The missing data could indicate that the raters ran out of time, and were unable to complete the task. Just by

looking at the data using this interpretation, the reliability does not seem to be too bad. There are two large differences in assessments: in unit 3 and especially in unit 9. In the other units, raters either agreed or chose neighbouring classes. The Frequency assessments (the odd-numbered units) show more disagreement than the Impact assessments (the even-numbered units). But this is just an approximate and informal appraisal. Computing alpha would give a more dependable valuation of the inter-rater reliability.

The first step in calculating alpha is to compute the observed coincidence matrix. Note that the score table contains 38 pairable scores; the rating for unit 14 cannot be paired and is therefore ignored. The o_{ck} matrix is defined by equation (A.2). For example:

$$o_{HM} = \frac{\overbrace{1}^{\text{unit 3}}}{3-1} + \frac{\overbrace{2}^{\text{unit 6}}}{3-1} + \frac{\overbrace{1}^{\text{unit 13}}}{2-1} = 2\frac{1}{2}$$

Completing this for all possible combinations gives (zero values omitted):

o_{ck}	k=U	L	M	H	V	Σ
c=U	1	$2\frac{1}{2}$		$\frac{1}{2}$		$4 = n_c$
L	$2\frac{1}{2}$	4	$3\frac{1}{2}$	1		11
M		$3\frac{1}{2}$	7	$2\frac{1}{2}$		13
H	$\frac{1}{2}$	1	$2\frac{1}{2}$	4	1	9
V				1		1
Σ	4	11	13	9	1	$38 = n$

The expected coincidence matrix was defined in equation (A.3). For example:

$$e_{HM} = \frac{9 \times 13}{38 - 1} = \frac{117}{37}$$

Completing this for all possible combinations gives:

e_{ck}	k=U	L	M	H	V	Σ
c=U	$\frac{12}{37}$	$\frac{44}{37}$	$\frac{52}{37}$	$\frac{36}{37}$	$\frac{4}{37}$	$4 = n_c$
L	$\frac{44}{37}$	$\frac{110}{37}$	$\frac{143}{37}$	$\frac{99}{37}$	$\frac{11}{37}$	11
M	$\frac{52}{37}$	$\frac{143}{37}$	$\frac{156}{37}$	$\frac{117}{37}$	$\frac{13}{37}$	13
H	$\frac{36}{37}$	$\frac{99}{37}$	$\frac{117}{37}$	$\frac{72}{37}$	$\frac{9}{37}$	9
V	$\frac{4}{37}$	$\frac{11}{37}$	$\frac{13}{37}$	$\frac{9}{37}$	0	1
Σ	4	11	13	9	1	$38 = n$

The distance metric δ^2 for ordinal data was defined in equation (A.4). As an example:

$$\delta_{UH}^2 = \left(\frac{4}{2} + (11 + 13) + \frac{9}{2} \right)^2 = (30\frac{1}{2})^2$$

Completing this for all possible combinations gives:

δ_{ck}^2	k=U	L	M	H	V
c=U	0	$(7^{1/2})^2$	$(19^{1/2})^2$	$(30^{1/2})^2$	$(35^{1/2})^2$
L	$(7^{1/2})^2$	0	12^2	23^2	28^2
M	$(19^{1/2})^2$	12^2	0	11^2	16^2
H	$(30^{1/2})^2$	23^2	11^2	0	5^2
V	$(35^{1/2})^2$	28^2	16^2	5^2	0

Using these matrices, equation (A.5) can be calculated. Because all matrices are symmetrical around their diagonal and because the diagonal of δ^2 is zero, the calculation can be restricted to the elements above the diagonal.

The result is $\alpha \approx 0.546$, which is considered a weak inter-rater reliability. Clearly, an even higher apparent similarity in assessments is needed in order to get $\alpha > 0.8$ that is considered sufficient for firm conclusions.

The informal appraisal also guessed that alpha for Frequency assessments (the odd-numbered units) would be less than that for Impact assessments. This can be verified by computing alpha over both subsets.

Frequencies	1	3	5	7	9	11	13
A	L	M	M	U	U	M	H
B	M	L	M	L	L	L	M
C	M	H	M	U	H	M	*

Variable o_{ck}^F denotes the observed coincidence for Frequency assessments only.

o_{ck}^F	k=U	L	M	H	V	Σ
c=U	1	$1/2$		$1/2$		$3 = n_c^F$
L	$2^{1/2}$		$2^{1/2}$	1		5
M		$2^{1/2}$	5	$1^{1/2}$		9
H	$1/2$	1	$1^{1/2}$		1	3
V						0
Σ	3	5	9	3	0	$20 = n^F$

Equation (A.7) computes to $\alpha^F \approx 0.289$. This is indeed much lower than the alpha over the entire set of units.

Doing the same for all even-numbered units (the Impact assessments):

Impacts	2	4	6	8	10	12	14
A	H	L	M	V	L	M	H
B	H	U	M	H	L	L	*
C	H	L	H	H	L	M	*

Similarly, o_{ck}^I denotes the observed coincidence for Impact assessments.

o_{ck}^I	k=U	L	M	H	V	Σ
c=U		1				$1 = n_c^I$
L	1	4	1			6
M		1	2	1		4
H			1	4	1	6
V				1		1
Σ	1	6	4	6	1	$18 = n^I$

Again using equation (A.7) α^I can be computed to $\alpha^I \approx 0.831$. The Impact-alpha is indeed much better than the Frequency-alpha.

Questionnaires

This Appendix describes the questionnaires that were used during the various lab and field experiments, and their results. It also contains the instructions that were provided for the reliability experiment in B.2.

B.1 Improving the initial design

Chapter 6 describes a case study in which the initial RASTER method is improved over a number of iterations. At the end of this case study, a questionnaire with two main questions was used. Each participant completed this questionnaire individually.

The next pages show a translation of this questionnaire, followed by the original questionnaire in Dutch.

Questionnaire

This questionnaire is about the diagrams that you created, and the risk analysis that you performed using them.

The questions concern risk factors. A risk factor is a general property that may influence the severity of a risk.

The following risk factors are used. For each, the right column shows a (fictitious) example.

A: Extent of negative consequences: damage to goods, and people's injuries.	<i>When an explosion occurs in this chemical factory, the material damage will amount to millions. Also, 200 people will have to be evacuated, and some will probably die from the immediate effects of the poison.</i>
B: Probability: how large the likelihood of negative consequences is, or how often they occur.	<i>However, such an explosion will occur only once every 10,000 years.</i>
C: Reversibility: the extent to which negative consequences can be repaired.	<i>The immediate surroundings will become permanently uninhabitable. In a nearby nature reserve an endangered species of butterfly will become extinct.</i>
D: Ubiquity: the size of the area in which negative consequences are noticeable.	<i>Also, farmers up to 150 km distance will have to destroy their crops out of precaution.</i>
E: Persistency: the duration of negative consequences.	<i>Cleaning up will take 5 years.</i>
F: Incertitude: doubt and uncertainty about the causes, the mechanisms, and consequences of the risk.	<i>A report was published that claimed that these figures are much too high. However, an other report indicates that there is little knowledge on the effects on children and people with existing health issues.</i>
G: Delayed effects: whether negative consequences occur immediately, or whether they only become apparent over time.	<i>Probably many effects on population health will only become apparent after many years.</i>

Question 1

The table on the next page shows all vulnerabilities that appear in your project. The columns of this table show the risk factors. Indicate, for each vulnerability, the relevance of each risk factor. Ask yourself the following question:

If I had to determine the severity of «vulnerability» for the crisis organisation of Radiocommunications Agency Netherlands, then the importance of «risk factor» in my considerations will be:

x → not important: its influence is limited, or the factor has no bearing whatsoever on the severity.

! → important: the factor has a significant or (very) large influence on the severity.

Write x or ! in each table cell, or leave the cell blank if you cannot determine the influence, or if you do not know the most appropriate answer.

The influence of an important factor can be increasing (aggravating) or decreasing (ameliorating). Both cases are indicated by the ! symbol.

Complete using: x = not relevant, ! = relevant, or leave the cell blank.

Type	Vulnerability	A:Extent of damage	B:Probability	C:Reversibility	D:Ubiquity	E:Persistency	F:Incertitude	G:Delayed effects
Equipment	Fire							
	Configuration							
	EMC							
	Physical impact							
	Natural disasters							
	Incorrect use							
	Flooding							
	Political tensions							
	SIM invalid							
	Power failure							
Wireless	Technical/human failure							
	Ageing							
	Attenuation							
	Congestion							
Wired	Jamming							
	Interference							
	Break							
	Congestion							
Ageing	EMC							
	Ageing							

Question 2

The next questions are each about the combination of one vulnerability and one component.

Question	Vulnerability	Component
2.a		
2.b		
2.c		
2.d		

For each of these combination you will give the best possible estimate of the value of each of the risk factors. You will do so by giving a lower bound and an upper bound.

Choose the highest and lowest values so that you are certain that the true value is somewhere between the lower and upper bound.

Make the bounds as narrow as possible, but make sure that you are fairly certain that the true value lies in between.

Note: in the actual questionnaire, specific vulnerabilities and components would have been filled in in the table above. Also, the next page would have been repeated four times, once each for questions 2.a to 2.d. As these pages are identical, only the one for question 2.a is shown overleaf.

Question 2.a

Vulnerability: _____

Component: _____

Factor	Lower and upper bounds to the size of the effect Effect on the telecom service
Extent of damage	Percentage unavailability of the service (100% is fully unavailable) at least _____ at most _____ _____ % _____ %
Probability	How often will the incident occur? For example: "once every 50 years" at least: once per _____ at most: once per _____ _____ year / month _____ year / month (mark)
Reversibility	Part of the effect that is permanent (0% = fully repairable, 100% = all effects are permanent) at least _____ at most _____ _____ % _____ %
Ubiquity	Size of the affected area, in relation to the field of action of the organisation (100% is entire field of action) at least _____ at most _____ _____ % _____ %
Persistency	Time required to restore the service entirely: at least: _____ at most: _____ minutes / hours / days minutes / hours / days weeks / months / years weeks / months / years (mark)
Incertitude	How certain are you about the correctness of the answers above? no doubt possible / very certain / fairly certain somewhat uncertain / very uncertain / don't know (mark)
Delayed effects	Time for the effects to become apparent at least: _____ at most: _____ minutes / hours / days minutes / hours / days weeks / months / years weeks / months / years (mark)

Vragenlijst

Deze vragenlijst gaat over de diagrammen die u hebt getekend, en de risicoanalyses die u daarbij hebt gemaakt.

In de vragen gaat het over risicofactoren. Een risicofactor is een algemene eigenschap die mogelijk de ernst van een risico beïnvloedt.

De volgende risicofactoren worden gebruikt. In de rechter kolom staat steeds een (fictief) voorbeeld ervan.

A: Grootte van nadelige gevolgen: de schade aan goederen, en verwonding van personen.	<i>Bij een ontploffing in deze chemische fabriek zal de materiële schade in de miljoenen lopen. Ook zullen 200 mensen geëvacueerd moeten worden, en zullen waarschijnlijk enkele mensen aan de directe gevolgen van het gif overlijden.</i>
B: Waarschijnlijkheid: hoe groot de kans op schadelijke effecten is, of hoe vaak die optreden.	<i>Zo'n ontploffing zal echter slechts eens in de 10.000 jaar optreden.</i>
C: Omkeerbaarheid: de mate waarin de schadelijke gevolgen hersteld kunnen worden.	<i>De directe omgeving zal permanent onbewoonbaar zijn. In het naastgelegen natuurgebied zal een zeldzame vlindersoort uitsterven.</i>
D: Uitgestrektheid: de grootte van het gebied waarin schadelijke effecten merkbaar zullen zijn.	<i>Ook zullen tuinders tot op 150km afstand hun oogst uit voorzorg moeten vernietigen.</i>
E: Langdurigheid: hoe lang de schadelijke effecten aanwezig blijven.	<i>De schoonmaakwerkzaamheden zullen 5 jaar duren.</i>
F: Onzekerheid: twijfel en onduidelijkheid over de oorzaken, de werking of de gevolgen van het risico.	<i>Er is een rapport uitgebracht waaruit zou blijken dat de genoemde cijfers veel te hoog zijn. Een ander rapport geeft echter aan dat er te weinig bekend is over de gevolgen op kinderen, of op mensen die al kampen met gezondheidsproblemen.</i>
G: Vertraagde effecten: of schadelijke effecten niet meteen zichtbaar zijn, maar sluimeren en pas na langere tijd merkbaar worden.	<i>Waarschijnlijk zullen veel gezondheidsproblemen pas jaren later aan het licht komen.</i>

Vraag 1

In de tabel op het volgende blad staan alle kwetsbaarheden die in uw onderzoek voorkomen. In de kolommen van de tabel staan de risicofactoren. Geef voor elke kwetsbaarheid aan wat de relevantie van ieder van deze risicofactoren is. Stel u daarbij de volgende vraag:

Als ik moet bepalen of (kwetsbaarheid) ernstig is of niet voor de crisisorganisatie van Agentschap Telecom, dan is (risicofactor) in mijn afweging

x → *niet belangrijk: de invloed ervan is beperkt, of de factor heeft geen enkele invloed op de ernst.*

! → *belangrijk: de factor heeft een behoorlijke tot (zeer) grote invloed op de ernst.*

Schrijf een **x** of **!** in het vakje, of laat het vakje op en als u de invloed niet kan bepalen, of als u het meest geschikte antwoord niet weet.

De invloed van een belangrijke factor kan verhogend (ernstiger) of juist verlagend (minder ernstig) zijn. Beide gevallen worden met een **!** aangegeven.

Vol in: **x** = niet belangrijk, **!** = belangrijk, of laat het vakje open.

Type	<i>Knetsbaarheid</i>	<i>A: Nadeltige gevolgen</i>	<i>B: Waarschijnlijkhed</i>	<i>C: Omkeerbaarheid</i>	<i>D: Uligestrekteheid</i>	<i>E: Langdurigheid</i>	<i>F: Onzekerheid</i>	<i>G: Verrangde effecten</i>
Apparatuur	Brand							
	Configuratie							
	EMC							
	Impact							
	Natuurrampen							
	Onjuist gebruik							
	Overstroming							
	Politieke spanning							
	SIM ongeldig							
	Stroomuitval							
	Technisch/menselijk falen							
	Veroudering							
	Afzwakking							
	Congestie							
	Jamming							
Draadloze verbinding	Storing							
	Breuk							
	Congestie							
	EMC							
Draadgebond en verbinding	Veroudering							

Vraag 2

De volgende vragen gaan telkens over een combinatie van een kwetsbaarheid en een component.

<i>Vraag</i>	<i>Kwetsbaarheid</i>	<i>Component</i>
2.a		
2.b		
2.c		
2.d		

U gaat voor elk van deze combinaties de best mogelijke schatting geven voor de waarde van elk van de risicofactoren. Dat doet u telkens door een laagste en hoogste waarde op te geven.

Kies de hoogste en laagste waarde zo dat u er zeker van bent dat de werkelijke waarde ergens tussen die ondergrens en bovengrens in ligt.

Maak de grenzen zo smal mogelijk, maar zorg ervoor dat u vrij zeker weet dat de juiste waarde zich daar tussen bevindt.

Vraag 2.a

	Onder- en bovengrens aan grootte van het effect
Factor	Effect op de telecomdienst(-verlening)
Grootte van nadelige gevolgen	uitvalpercentage van de dienst (100% is volledige uitval) minimaal maximaal %
Waarschijnlijkheid	Hoe vaak zal het incident voor komen? Voorbeeld "eens per 50 jaar". minimaal: eens per maximaal: eens per maand / jaar maand / jaar (omcirkel)
Omkeerbaarheid	Deel van de effecten die permanent zijn (0% = volledig herstelbaar, 100% = alle effecten zijn permanent) minimaal maximaal %
Uitgestrektheid	Grootte van het getroffen gebied, in verhouding tot het werkgebied van de organisatie (100% is gehele werkgebied) minimaal maximaal %
Langdurigheid	Tijd voordat de dienst geheel hersteld is: minimaal maximaal minuut/uur/dag/ week/maand/jaar minuut/uur/dag/ week/maand/jaar (omcirkel)
Onzekerheid	Hoe zeker bent u van de juistheid van bovenstaande antwoorden? geen twijfel mogelijk – heel zeker – vrij zeker – enigszins onzeker – erg onzeker – geen idee (omcirkel)
Vertraagde effecten	Tijd voordat effecten zichtbaar worden is minimaal maximaal minuut/uur/dag/ week/maand/jaar minuut/uur/dag/ week/maand/jaar (omcirkel)

B.2 Reliability experiment

Chapter 7 describes an experiment in which groups of students were to perform part of the RASTER project. The next pages show a transcription of the handouts that were provided to the students. These handouts describe a fictitious company for which the participants were to assess likelihood and impact of various hazards to components.

The handouts are followed by the original exit questionnaire for this experiment. The results of this questionnaire were shown in Figure 7.4.

About Fold Factory

Fold Factory is a small company that creates luxury packing cartons. These are not average boxes, but artistic creations made specifically on request of customers. Below are some examples of their work.

The box on the left was created for an organic farm in Sweden. The box is made from a single piece of carton, and is entirely recyclable. One important design criterion was that the box should be able to firmly hold eggs of non-uniform sizes.

The box on the right was created for a traditional butchery in Spain, who wanted to start selling their products to consumers directly, using an online shop. This design won the prestigious Food and Wine Industry Award for best packaging in 2011.

Fold Factory was created by two friends, at a time when they were both students at a school of Industrial Design. The company quickly grew to its current size of about 15 employees. Fold Factory designs the packing concepts, the packaging itself and the artwork; the actual printing and cutting are outsourced to specialist printers throughout Europe. The design department consists of 8 creative artists and packaging specialists. They use high-powered workstations for all their art work and 3D modelling. Other departments handle marketing, sales, and contacts with printers. There is a small support staff.

Because of their mode of doing business, e-mail is very important to Fold Factory. E-mail is used to communicate with prospective customers, existing customers, and printers. Of course, the telephone is an important alternative (especially for customers), but e-mail is necessary for transmission of file attachments. Artwork proposals are sent to customers, and cutting models are sent to printers.

Fold Factory relies heavily on IT. The IT infrastructure is owned by the company itself. Most users are computer savvy, and there is little need for support on PCs, laptops, and workstations. The servers, however, are maintained by a local IT shop. All servers and networking equipment are stored in a single room. This room is locked and air-conditioned.

Fold Factory is housed on the second floor of a building on a large office complex in the Amsterdam area. The other floors are occupied by accounting and consulting companies.

Diagram

Components

Department 1, 2: an ethernet network that only contains PCs (desktops, laptops, workstations). No servers are connected to these networks.

desk cable: a plain ethernet cable, used to connect a user's laptop, desktop or workstation to the nearest wall outlet.

desktop: a simple PC. Mainly used by office workers.

DHCP server: an old machine that is currently only used to provide network configuration (IP addresses, etc) to desktops, laptops, and workstations.

DMZ ("demilitarized zone"): a small network that only contains servers. There are never direct connections between computers on the internal network and the internet; there are only connections from internet to DMZ and from DMZ to the internal network.

DNS server: provides name to IP address lookups for all computers in the company; used for intranet addresses as well as any other domain name lookups.

ethernet cable: normal ethernet network cables.

external mail server: any mail server for a customer, printer, or any other external contact.

fiber optic: a fiber optic cable, provided from some Internet Service Provider, connecting Fold Factory to the internet.

File server: central storage of documents, artwork, box designs, etc. Some company confidential information is stored here, as well as confidential financial and customer information. The file server is regularly backed up by the local IT support company.

Firewall DMZ: controls the flows of information between the internal network and the DMZ ("demilitarized zone"). There are never direct connections between computers on the internal network and the internet; there are only connections from internet to DMZ and from DMZ to the internal network.

Firewall ext(ernal): a firewall that also acts as router between the internet and Fold Factory's network.

Firewall int(ernal): separates the server LAN from the PCs, but also acts as a network switch between the departments and the server LAN.

internet: the Internet Service Provider for Fold Factory, and the rest of internet.

laptop: a laptop PC. Mainly used by traveling sales staff and managers.

Mail server: all email to and from computers is sent through the internal mail server. All past email is stored here as well.

MX 1, 2: external mail servers, maintained by the Internet Service Provider, that temporarily hold incoming email for Fold Factory when the relay server is unreachable. For example, when the internet connection to Fold Factory is down, email to Fold Factory is kept on the MX servers. MX 1 and 2 are maintained by the Internet Service Provider.

relay server: a mail server. It receives outgoing email from the Mail server, and passes it on to any external mail server on the internet. It also receives incoming email from the internet, performs spam filtering, and passes email on to the internal Mail server.

Server LAN: an ethernet network that only contains servers. No PCs are connected to this network.

work station: a powerful PC, used to create artwork and 3D models. Used by the creative artists and engineering teams.

Vulnerabilities

The following vulnerabilities are used for the e-mail service of Fold Factory.

For cables

Break:	Cable is damaged, cut, or disconnected by natural events, digging of a trench, accidental unplugging, improper maintenance, twisting, bending, or some other external influence.
Congestion:	The amount of traffic offered exceeds the capacity of the link.
Cable ageing:	Damage caused by passing of time. The outer insulation of the cable may weaken, and joints and connectors may get corroded.

For equipment items

Physical damage:	Fire, burns, floods, spills, water from fire extinguishing installations, knocks, falls, drops, kicks, and other physical damage.
Power:	Failure of the electrical power supply. Power surges, or fluctuations in voltage.
Configuration:	Incorrect configuration settings (in hardware or software) or mistakes made by system administrators, operators, or end users.
Equipment ageing:	Failures caused by passing of time. Mechanical parts such as hard disks and fans wear out, insulation weakens, etc.
Theft:	Malicious stealing of parts or physical equipment by insiders or outsiders.
Information:	Missing or inconsistent data, affecting the operations of the equipment. This can be through human error, equipment failure, malicious tampering, hacking, or any other natural or man-made cause.

Estimating frequency

The factor Frequency indicates the likelihood that the vulnerability will lead to an incident.

<i>Class</i>	<i>Value</i>	<i>Symbol</i>
Extremely low	Very rare, but not physically impossible.	U
Low	Once in 500 years, or: for 1000 identical components, each year 2 will experience an incident.	L
Medium	Once in 50 years, or: for 1000 identical components, each month 1 or 2 will experience an incident.	M
High	Once in 5 years, or: for 1000 identical components, each month 15 will experience an incident.	H
Extremely high	Routine event. Very often.	V
Unknown	Indicates lack of knowledge or data, or lack of consensus.	?

A frequency of “once in 50 years” is an average, and does not mean that each 50 years an incident is guaranteed to occur. It may be interpreted as:

- The average timespan between incidents on a single component is 50 years.
- For a set of 50 identical components, each year on average one of them will experience an incident.
- Each year, the component has a 1 in 50 chance of experiencing an incident.

When the life time of a component is 5 years (or when the component is replaced every 5 years) the frequency of a vulnerability can still be “once in 500 years”.

Example: a component is always replaced after one year, even if it is still functioning. On average, 10% of components fail before their full year is up. The general frequency for this failure is therefore estimated as “once in 10 years” even though no component will be in use that long.

Note that this value is between the characteristic values for High and Medium. The team members must together decide which of these two classes is assigned.

Procedure to determine Frequency

Use the following four-step procedure to determine the factor Frequency:

1. Find the frequency class that applies to this type of node in general.
This can be based on, for example, past experience or expert opinion.
2. Think of reasons why this particular node should have a lower frequency than usual.
Existing countermeasures may make the frequency lower than usual. For example, if an organisation already has a stand-by generator that kicks in when power fails, then the frequency of power failure incidents is thereby reduced. The frequency does not reflect the likelihood that the vulnerability is triggered, but the likelihood that the vulnerability will lead to a noticeable incident. Another example is the use of premium quality components, or secure and controlled equipment rooms.
For some components monitoring can detect failures that are imminent before they occur. This also will reduce the frequency of incidents.
3. Think of reasons why this particular node should have a higher frequency than usual.
The environment may be particularly dangerous, for example, for emergency services such as fire fighting teams.
4. Decide on the frequency class for this particular node.
Typically either Low, Medium, or High will be used.
If neither of these accurately reflect the frequency, one of the extreme classes should be used.
If no class can be assigned by consensus, Unknown should be used.

Estimating impact

The factor Impact indicates the severity of the effect when a vulnerability does lead to an incident. This severity is the effect to the service as a whole, not its effect to the component that experienced the vulnerability. For example, a power failure will cause equipment to stop functioning temporarily. This is normal, and of little relevance. What is relevant is the effect to the service. The outage could cause the service to fail (if the equipment is essential), but could also have a no effect at all (if the equipment has a backup). Or any effect in between.

Only the effects on the telecom service must be taken into account in this stage. Loss of business, penalties, and other damage are not considered, but may be relevant during later stages in the process (not part of this experiment).

The damage may be caused by an incident that also affects other components. For example, a cable may be damaged by an earthquake; the same earthquake will likely cause damage to other components as well. However, this additional damage must not be taken into account. Only the damage resulting from the damage to this component must be considered. An other stage in the process (not part of this experiment) takes care of multiple failures due to a single incident.

For the Impact, the possible values are:

<i>Class</i>	<i>Value</i>	<i>Symbol</i>
Extremely low	Unnoticeable effects.	U
Low	Noticeable degradation of the service.	L
Medium	Partial temporary unavailability of the service for some actors.	M
High	Long-term, but eventually repairable unavailability of the service for all actors.	H
Extremely high	Very long-term or unrepairable unavailability of the service for all actors. Major redesign of the telecom service is necessary, or the service has to be terminated and replaced with an alternative.	V
Unknown	Indicates lack of knowledge or data, or lack of consensus.	?

The expected time to repair must also be taken into account. Repair time is expressed in the phrases “long-term”, “temporary”, and “unrepairable”.

Procedure to determine Impact

Use the following four-step procedure to determine the factor Impact:

1. Choose the impact class that most accurately seems to describe the impact of the incident.
2. Think of reasons why the impact would be higher than this initial assessment.
3. Think of reasons why the impact would be lower than the initial assessment.
Existing redundancy can reduce or even annul the impact. For example, a telecom service may have been designed such that when a wireless link fails, a backup wired link is used automatically. The impact of the wireless link failing is thereby reduced. Monitoring and automatic alarms may reduce the impact of incidents. When incidents are detected quickly, repairs can be initiated faster.
Keeping stock of spare parts, well trained repair teams, and conducting regular drills and exercises all help in reducing the impact of failures and must be considered in the assessment.
4. Decide on the impact class.

Typically either Low, Medium, or High will be used. If neither of these accurately reflect the impact, one of the extreme classes should be used. If no class can be assigned by consensus, Unknown should be used.

Remember

- Refer to these instructions and tables often, and use them to guide your estimates.
- Work as a team, and convince each other with good arguments.
- Be consistent. Review your estimates to ensure that like cases are scored alike.
- Mark your answers (and corrections) as in the example below.

Keep remarks for your own use, if you wish

Node name	Vulnerability	Frequency	Impact	Your notes
mobile phone (equipment)	Dropped on floor	U - L - M - H - V - ?	U - L - M - H - V - ?	Phone can handle most falls...
	Stolen	U - L - M - H - V - ?	U - L - M - H - V - ?	
	Water damage	U - L - M - H - V - ?	U - L - M - H - V - ?	Use protective cover!

Make corrections like this

For the recorder

- Help the group to follow the instructions for making estimates. When estimates are becoming easier, attention may slip. It is still important to follow the proper procedure.
- Watch out for inconsistencies in the the estimates by your group.
For example, when two vulnerabilities have comparable frequency, the group should not score one as “Low” and the other as “Medium”; both should be either “Low” or both “Medium”. Bring inconsistent scores to the attention of other group members.
- If the group makes important assumptions, make sure to mark them in the “Your notes” column.
- Maintain the group scores, and hand those in at the end of the experiment.

Note: In the original handouts now followed the scoring forms. These forms are not included here for brevity. The forms contained one entry for each node in the diagram, as per the example above.

The vulnerabilities on nodes were as described previously. For cables: break, congestion and cable ageing. For equipment items: physical damage, power, configuration, equipment ageing, theft, and information.

Reliability experiment questionnaire

Each participant individually completed an exit questionnaire at the end of the experiment. The goal of the questionnaire is to check whether mitigations of contextual sources of variation have been successful. For each countermeasure the questionnaire checks whether participants have the required knowledge, ability and motivation to apply the mitigation measure. For example, for 'lack of knowledge or experience' we use these three questions: "My knowledge of the technology behind office email services can be described as (non-existent – excellent)" (knowledge), "My knowledge of the technology behind office email services could be applied in the exercise" (ability), and "It was important that my knowledge of email services was used by the group" (motivation).

We also used the opportunity to include four questions to test some internal sources of variation. In particular, we wanted to test whether the scales defined for Frequency and Impact were suitable, and whether the procedure to avoid intuitive and potentially biased assessments was effective.

To encourage honest answers and prevent participants from giving the 'desired' answer, the order of the questions is shuffled. Also the scales for some of the questions are reversed, so that 'good' answers do not always belong to the right-most column of the exit questionnaire form.

The next pages show the original questionnaire.

Exit questionnaire

This questionnaire will take approximately 5 minutes to complete.

The topic of the questionnaire is the exercise that you just completed: the introduction and explanations at the start, your application of the instructions, and the estimates made by you and your group.

The questionnaire consists of 23 statements.

For each statement, circle the column that most accurately describes your opinion on the matter, or best completes the sentence.

If you make a mistake, correct it by crossing out the wrong answer and marking the right one.

Examples:

My team members and I worked on the exercise in a cooperative and pleasant manner.	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Discussions in my team were ...	very short	short	just right	long	very long

Now, please consider each of the following 23 statements.

During the experiment I knew how to avoid fast, intuitive estimates.	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
The instructions that were given at the start of the exercise were ...	very useful	somewhat useful	neither useful nor useless	somewhat useless	mostly useless
I knew what I needed to do during the exercise.	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
I was able to concentrate on the exercise and work comfortably.	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Participating in this experiment was ...	very interesting	interesting	neither interesting nor tiresome	somewhat tiresome	very tiresome
My knowledge of the technology behind office email services can be described as ...	non-existent	very limited	neither good nor limited	good	excellent
The scales for values of Frequency and Impact estimates were ...	very clear	clear	neither clear nor unclear	unclear	very unclear
The final answer of my group often equalled my immediate personal estimate.	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
In my group we mostly had the same ideas on the values of estimates.	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
My knowledge of the technology behind office email services could be applied in the exercise.	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
When estimating Frequencies and Impacts of vulnerabilities, it is necessary to consider many possible incidents.	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
For all estimates, there exists a single best value (whether we identified it or not).	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
It was important that my knowledge of email services was used by the group.	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
The instructions at the start of the exercise were ...	very clear	somewhat clear	neither clear nor unclear	somewhat unclear	very unclear

Please continue on next page →

The estimates made by other groups (compared to ours) will be ...	very similar	somewhat similar	neither similar nor different	somewhat different	very different
In the experiment I could practically apply the instructions that were given at the start of the exercise.	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
The instructions and procedures for avoiding fast, intuitive estimates were ...	very cumbersome	somewhat cumbersome	neither easy nor cumbersome	somewhat easy to use	very easy to use
I could think of practical examples for most of the vulnerabilities.	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Before the exercise I was instructed to make rational, calculated estimates.	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
In my group we hesitated between two adjacent Frequency and Impact values ...	almost always	often	neither seldom nor often	seldom	almost never
The scales of values for Frequency and Impact were suitable to this exercise.	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
When discussing vulnerabilities, other members of my group often gave examples that I would never have thought of.	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
The time to complete the exercise was ...	more than sufficient	sufficient	just right	somewhat short	way too short

My group number was: _____

I was recorder for this group: Yes / No

Thank you!

B.3 Field test questionnaires

Chapters 8 and 9 describe two field tests to validate RASTER in practice. After each field tests the participants completed a questionnaire.

Most questions were phrased as a statement; participants indicated their agreement with the statement using a five-point scale (strongly disagree to strongly agree). Some questions were expressed as a sentence to be completed; participants chose the words that formed the best completion. For example, the question “The explanations and instructions on the RASTER method were . . .” could be completed by choosing one of: very unclear, unclear, neither unclear nor clear, clear, very clear. The order of the questions was randomised, and for some of the questions the natural order of the possible answer (most negative to most positive) was reversed (most positive to most negative). This was done to stimulate careful completion of the questionnaire.

The table below shows the results, with the questions sorted and grouped by topic. The marker ● indicates scores for the first field test; each marker indicates a single participant’s answer. The marker ▲ represents the second field test. Above those scores, codes SD, D, N, A, SA stand for strongly disagree, disagree, neither agree nor disagree, agree, strongly agree respectively. Otherwise, the codes are clarified in the text in the first column.

<i>Ability to participate</i>					
1. The explanations and instructions on the RASTER method was (very unclear . . . very clear).	VU	U	N	C ●●●●● ▲▲▲▲▲	VC ●● ▲
2. The explanations and instructions on the goals of the project were (very unclear . . . very clear).	VU	U	N ▲	C ●●●●● ▲▲▲▲▲	VC ● ▲
3. I knew what input was expected from me during the project.	SD	D ▲	N ● ▲	A ●●●●● ▲▲	SA ▲
4. I was able to apply the explanations and instructions on the RASTER method.	SD	D	N ●● ▲▲	A ●●●●● ▲▲▲▲▲	SA
<i>Ability to contribute</i>					
5. My technical knowledge on telecom services is (very limited . . . excellent).	VL	Limtd. ●● ▲▲	N ●● ▲▲	Good ●●● ▲	Exc.
6. My knowledge on the use and purpose of telecom services in practice is (very limited . . . excellent).	VL	Limtd. ●	N ●● ▲▲▲	Good ●●●●● ▲▲	Exc.
7. With my knowledge I was able to contribute to the project.	SD	D	N ●● ▲▲	A ●●●●● ▲▲▲▲▲	SA ●

Continued on next page

Attitude to consensus

8. In the project group we agreed on the estimates for Frequency and Impact.	SD	D	N ● ▲	A ●●●●● ▲▲▲	SA ● ▲
9. In principle there is always one best estimate (whether we always found it or not).	SD	D	N ●●●●● ▲▲▲	A ▲▲	SA ●

Clarity of the scales

10. The scale used for Frequency was (very unclear ... very clear).	VU	U	N ▲	C ●●●●● ▲▲▲▲	VC
11. The scale used for Impact was (very unclear ... very clear).	VU	U	N	C ●●●●● ▲▲▲▲▲	VC

Applicability of the scales

12. When making estimates on Impacts I ... hesitated between two neighbouring classes (almost always ... very seldom).	AA	Often ● ▲▲▲▲	N ●● ▲	Seldom ●●●●	VS
13. When making estimates on Frequencies I ... hesitated between two neighbouring classes (almost always ... very seldom).	AA	Often ● ▲▲	N ▲▲▲	Seldom ●●●●●	VS
14. The scale used for Frequency was suitable for this project.	SD	D ● ▲▲	N	A ●●●●● ▲▲▲	SA
15. The scale used for Impact was suitable for this project.	SD	D ▲	N ▲	A ●●●●● ▲▲	SA
16. The group's estimated matched my personal estimates.	SD	D ▲	N ●●● ▲▲	A ●●● ▲▲	SA ●

Time required by the analysts

17. The time available for this project was (very tight ... more than ample).	VT	Tight ●●	N ●●●● ▲▲▲▲	Ample ▲	MTA
18. If we had spent more time on the project, the results would have been (much worse ... much better).	MW	W	N ●●●●● ▲▲▲▲	B ● ▲	MB
19. It was easy to make time available for this project.	SD	D ●●●● ▲▲▲	N ● ▲	A ●● ▲	SA
20. If the project had required more time, I would not have been able to participate.	SA ▲	A ●	N ●●●	D ●●● ▲▲▲	SD ▲
21. The project took a lot of my time.	SA	A	N ●●●● ▲▲▲	D ●●● ▲▲	SD

Continued on next page

Attitude to the project

22. Participating in this project was (very boring ... very interesting).	VB	B	N	I ●●● ▲▲	VI ●●● ▲▲▲
---	----	---	---	----------------	------------------

Need for localisation

23. I found it difficult that the RASTER manual was only available in English.	SD	D ●●	N ●●	A ●●●	SA
24. It is important that all RASTER documentation is available in Dutch.	SD	D ● ▲	N	A ●●●● ▲▲▲	SA ● ▲

Accuracy and efficiency of risk identification

25. The RASTER method assists in quickly determining all large risks.	SD	D ●	N ▲▲	A ●●●● ▲▲▲	SA ●
26. There exist large risks for unavailability of telecom services that were not found in this project.	SA	A ●	N ●●●● ▲	D ● ▲	SD ● ▲▲
27. The RASTER method helps to ignore all small risks quickly.	SD	D ● ▲▲	N ▲▲▲	A ●●●●●	SA
28. The final risk list also contains risks that actually are not that important.	SA	A ● ▲	N ● ▲	D ●●●● ▲▲▲	SD ●

Correctness of the outcome

29. I am confident that the results of the project are correct.	SD	D	N ●	A ●●●● ▲▲▲▲	SA ●● ▲
30. I can take responsibility for the results of the project towards my colleagues.	SD ▲	D	N	A ●●●● ▲▲	SA ●●● ▲▲▲

Correctness of risk prioritisation

31. The RASTER method helps to determine risk priorities.	SD ●	D	N	A ●●●● ▲▲▲▲▲	SA ●●
32. The risks on the final risk list are shown in the right order (highest priority first).	SD	D	N	A ●●●●●● ▲▲▲▲▲	SA

Effectiveness of the project

33. I am surprised by the results of this project.	SD ●	D ●● ▲▲	N ●●● ▲▲▲	A ●	SA
34. Altogether, the project has made [organisation] (much less secure ... much more secure).	MLS	Less S.	N ●●●● ▲▲▲	More S. ●●● ▲▲	MMS

Continued on next page

Clarity of procedures

35. For me drawing and discussing the telecom diagrams was (very unnecessary ... very useful).	VU	Somewhat Unnry.	N	Somewhat Usefl. ●	VU ●●●●● ▲▲▲▲▲
36. I found the telecom diagrams (very unclear ... very clear).	VU	U	N ●	C ● ▲▲▲▲	VC ●● ▲
37. For me estimating and discussing Frequency and Impact was (very unnecessary ... very useful).	VU	Somewhat Unnry.	N ●●	Somewhat Usefl. ▲	VU ●●●●● ▲▲▲▲
38. I found the estimates of Frequency and Impact (very unclear ... very clear).	VU	U	N ▲	C ● ●●●●● ▲▲▲▲▲	VC

Technology Acceptance Model – Perceived Ease of Use

39. Learning to operate RASTER would be easy for me.	SD	D ▲	N ● ▲▲	A ●●●●● ▲▲	SA
40. I would find it easy to get RASTER to do what I want it to do.	SD	D ● ▲▲▲	N ●● ▲▲	A ●●●	SA
41. My interaction with RASTER would be clear and understandable.	SD	D ● ▲	N ● ▲▲	A ●●●●● ▲▲	SA
42. I would find RASTER to be flexible to interact with.	SD	D	N ●●●● ▲▲▲	A ●● ▲▲	SA ●
43. It would be easy for me to become skilful at using RASTER.	SD	D ● ▲	N ▲	A ●●●●● ▲▲▲	SA
44. I would find RASTER easy to use.	SD	D ▲▲	N ●●● ▲▲	A ●●●●● ▲	SA

Technology Acceptance Model – Perceived Usefulness

45. Using RASTER in my job would enable me to accomplish tasks more quickly.	SD	D ●● ▲▲	N ▲	A ●●●●● ▲▲	SA ●
46. Using RASTER would improve my job performance.	SD	D ▲	N ▲▲	A ●●●●● ▲▲	SA ●●●
47. Using RASTER in my job would increase my productivity.	SD	D ● ▲	N ●●● ▲▲	A ●● ▲▲	SA ●
48. Using RASTER would enhance my effectiveness on the job.	SD	D	N ●● ▲▲	A ●●●●● ▲▲	SA ▲
49. Using RASTER would make it easier to do my job.	SD	D ▲	N ▲	A ●●●●● ▲▲▲	SA ●
50. I would find RASTER useful in my job.	SD	D ▲	N ●● ▲	A ●●●●● ▲▲▲	SA ●

The next pages show the original questionnaire. Note that this questionnaire was used within Dutch crisis organisations; participants strongly preferred to receive all materials in Dutch. A translation of the questions and their possible answers was already provided in the table on the previous pages.

Deelnemer: _____

Afsluitende vragenlijst

Deze vragenlijst neemt ongeveer 20 minuten in beslag.

Het onderwerp van deze vragenlijst is het gehele project dat bij de Veiligheidsregio Groningen is uitgevoerd: de introductie, werkwijze, uitvoering en resultaten.

In het project hebben we gebruik gemaakt van Raster, een methode om een risico-beoordeling voor uitval van telecomdiensten te maken.

De vragenlijst bevat 50 stellingen.

Voor elke stelling omcirkelt u het antwoord dat het beste overeenkomt met uw persoonlijke mening, of het antwoord dat de zin naar uw mening het beste afmaakt.

Voorbeelden

Er was in voldoende mate koffie aanwezig tijdens de bijeenkomsten.	Helemaal mee oneens	Mee oneens	Neutraal	<u>Mee eens</u>	Helemaal mee eens
De koffie was ...	veel te heet	iets te heet	<u>precies goed</u>	iets te koud	veel te koud

Met deze vragenlijst kunnen wij beter begrijpen wat in de praktijk de voor- en nadelen zijn van de Raster methode.

De vragenlijst start op de volgende bladzijde.

Ik heb vertrouwen in de juistheid van de uitkomsten van het project.	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens
Als ik zelf een telecom-risicoanalyse moest doen, dan zou de Raster methode mijn uitvoering verbeteren.	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens
Er is in principe altijd één beste schatting (of we die nu altijd vonden of niet).	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens
De beschikbare tijd om dit project uit te voeren was ...	meer dan genoeg	ruim	voldoende	krap	veel te krap
Mijn technische kennis van telecomdiensten is ...	zeer beperkt	beperkt	goed noch beperkt	goed	uitstekend
Het zou voor mij helder en begrijpelijk zijn om zelf met de Raster methode te werken.	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens
In de projectgroep waren wij het eens over de schattingen voor Waarschijnlijkheid en Impact.	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens
De gebruikte schaal voor Waarschijnlijkheid was ...	zeer helder	helder	helder noch onduidelijk	onduidelijk	zeer onduidelijk
Als ik zelf een telecom-risicoanalyse moest doen, dan zou de Raster methode mijn productiviteit verhogen.	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens
De gebruikte schaal voor Waarschijnlijkheid was geschikt voor dit project.	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens
Ik vond de schattingen van Waarschijnlijkheid en impact ...	zeer duidelijk	duidelijk	duidelijk noch onduidelijk	onduidelijk	zeer onduidelijk
Ik zou de Raster methode eenvoudig vinden om zelf te gebruiken.	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens
De Raster methode helpt om alle grote risico's snel te vinden.	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens
Als ik zelf een telecom-risicoanalyse moest doen, dan zou de Raster methode mij in staat stellen om taken sneller uit te voeren.	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens
Bij het schatten van Impact twijfelde ik ... tussen twee naastgelegen waarden.	bijna altijd	vaak	vaak noch zelden	zelden	zeer zelden
Deelnemen aan dit project was ...	zeer interessant	interessant	interessant noch saai	enigszins saai	zeer saai
Ik durf verantwoording te nemen voor de resultaten van het project tegenover mijn collega's.	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens
Al met al heeft dit project ertoe geleid dat de VRG ... is geworden.	veel onveiliger	onveiliger	veiliger noch onveiliger	veiliger	veel veiliger
Als wij meer tijd aan het project hadden besteed, dan zouden de resultaten ... zijn geweest.	veel slechter	minder goed	hetzelfde	beter	veel beter
Bij het schatten van Waarschijnlijkheid twijfelde ik ... tussen twee naastgelegen waarden.	bijna altijd	vaak	vaak noch zelden	zelden	zeer zelden

Naar volgende pagina →

Ik zou het eenvoudig vinden om zelf de Raster methode te laten doen wat ik wil doen.	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens
Ik vond de telecomdiagrammen ...	zeer duidelijk	duidelijk	duidelijk noch onduidelijk	onduidelijk	zeer onduidelijk
Ik zou zelf de Raster methode flexibel vinden in het gebruik.	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens
Mijn kennis over het gebruik en nut van telecomdiensten in de praktijk is ...	zeer beperkt	beperkt	noch goed noch beperkt	goed	uitstekend
Ik wist welke inbreng er tijdens het project van mij verwacht werd.	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens
De uitleg en instructies over de Raster methode was ...	zeer duidelijk	duidelijk	duidelijk noch onduidelijk	onduidelijk	zeer onduidelijk
De gebruikte schaal voor Impact was geschikt voor dit project.	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens
Het zou voor mij eenvoudig zijn om de Raster methode zelf te leren gebruiken.	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens
Ik vond het lastig dat het Raster boekje alleen in het Engels beschikbaar was.	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens
Ik vond het maken en bespreken van schattingen van Waarschijnlijkheid en Impact ...	zeer nuttig	een beetje nuttig	nuttig noch overbodig	een beetje overbodig	zeer overbodig
De schattingen van de groep kwamen overeen met mijn persoonlijke schatting.	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens
Ik vond het tekenen en bespreken van de telecomdiagrammen ...	zeer nuttig	een beetje nuttig	nuttig noch overbodig	een beetje overbodig	zeer overbodig
Er bestaan grote risico's voor de uitval van telecomdiensten die in dit project niet gevonden zijn.	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens
Ik ben verrast door de uitkomsten van dit project.	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens
De uitleg over de opzet en doelen van het project was ...	zeer duidelijk	duidelijk	duidelijk noch onduidelijk	onduidelijk	zeer onduidelijk
De Raster methode helpt om de prioriteit van risico's te bepalen.	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens
Ik heb de uitleg en instructies over de Raster methode kunnen toepassen.	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens
Het is belangrijk dat alle Raster documentatie in het Nederlands is gesteld.	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens
Als ik zelf een telecom-risicoanalyse moest doen, dan zou de Raster methode mijn effectiviteit verhogen.	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens
Het was eenvoudig om tijd vrij te maken voor dit project.	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens

Naar volgende pagina →

De gebruikte schaal voor Impact was ...	zeer helder	helder	helder noch onduidelijk	onduidelijk	zeer onduidelijk
Ik kon met mijn kennis bijdragen aan het project.	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens
In de uiteindelijke risicolijst staan alle risico's in de juiste volgorde (hoogste prioriteit bovenaan).	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens
Op de uiteindelijke risicolijst staan ook risico's die eigenlijk niet zo belangrijk zijn.	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens
Als het project meer tijd gevraagd zou hebben, dan zou ik niet mee hebben kunnen doen.	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens
Ik zou de Raster methode nuttig vinden in mijn werk.	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens
Het zou voor mij eenvoudig zijn om vaardigheid te verkrijgen in de Raster methode.	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens
Als ik zelf een telecom-risicoanalyse moest doen, dan zou de Raster methode dat vergemakkelijken.	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens
De Raster methode helpt om alle kleine risico's snel te negeren.	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens
Ik was veel tijd kwijt aan de uitvoering van dit project.	helemaal mee oneens	mee oneens	neutraal	mee eens	helemaal mee eens

Hartelijk dank voor uw medewerking.

Raster software tool

To assist in drawing telecom service diagrams and to facilitate record-keeping when executing a RASTER project, the RASTER tool was developed. Note that the RASTER tool is not essential; it is a tool that supports execution of the method. The tool makes it easier to edit diagrams and to record and compare the assessments, but in principle the RASTER method can be performed independently of the tool. In fact, some of the experiments have been performed using pen and paper. However, during the course of further improvements to both the tool and the method it has become more difficult to maintain this strict separation. Executing the Raster method without any tool support has become practically impossible. The RASTER tool was not subject of this research project; testing the usability of the tool was not a goal of the validation described in the previous chapters.

The RASTER tool is web-browser-based, and is built using simple and widely-used web technologies: HTML5, CSS2, and Javascript on the client side, and PHP5 on the server side. The server-side PHP scripts are only used to store,

Figure C.1 – A screen shot of the RASTER tool in action.

Drawing diagrams	Adding and removing components, annotations, colours, connections. Ensuring consistent use of node classes and identical nodes, within and across diagrams.
Record keeping	Recording assessments, computing risk levels.
Consistency	Facilitating reviews to assist consistent scoring ("like assessments for like vulnerabilities").
Reports	Creating overviews and reports for use during analysis. At a glance overview of the most risky components. Longlist.
Collaboration	Facilitating discourse between technical and non-technical experts through rich graphical models. Assists reasoning about telecom service architectures.
Data collection	Research support by collecting statistics on components and assessments.
Project administration	Storing and retrieving projects; saving and printing projects.

Figure C.2 – Functions of the RASTER tool.

list, and retrieve projects on the server for collaboration; all manipulation is performed on the client side. The user manual for the tool is part of the RASTER Application Manual (see Appendix D).

As an indication of the size of this project, at the time of writing the tool consists of 10,219 lines of Javascript, excluding standard libraries. Localisation consists of 349 natural language expressions, currently in English and Dutch. Most of the HTML is generated by Javascript; 323 CSS rules take up 1,301 lines of text, both for on-screen display and for print media.

The entire software is available as free software under the GNU General Public License version 2 from Github. The URL is

<https://github.com/EelcoV/RasterTool/>.

Raster Method – application manual

This appendix contains a copy of selected chapters of the English RASTER manual. The manual is used by analysts when performing a RASTER analysis. It also serves as a reference manual for the RASTER tool, although only the analyst acting in the role of Recorder may actually use the software. The chapters included here (1 to 7) describe the method itself; chapters 8 to 13 documenting the RASTER tool have been excluded. English and Dutch versions of the full manual are available as a free download, in print, on-screen and e-book versions. For the URL see Appendix C.

In this reproduction layout styles have been adjusted to fit the rest of this thesis, and colours have been converted to greys. Other than those visual changes, this appendix is a verbatim reproduction of the RASTER manual. Unfortunately, this also means that this appendix uses its own section numbering.

1 Introduction

Introduction and guide to this document.

1.1 The problem

Organisations use many types of telecommunication services: fixed and mobile telephony, videoconferencing, internet, encrypted links between offices, etc. In the last decade, organisations have become much more dependent on these services. Whereas in the past a telephone outage was an inconvenience, today the failure of telecom services often makes it impossible to do business at all. And as organisations move online and into the cloud, reliability of telecom services becomes even more essential.

At the same time, technological and market changes have made it more difficult to assess the reliability of telecommunication services. Networks grow continuously, new technologies replace old ones, and telecom operators outsource and merge their operations. For any end-to-end telecom service, several telecoms operators will be involved, and none of them can understand how important that service is to each customer.

This increased dependency applies even more to crisis organisations (police, fire services, medical care, etc) and for crisis support and decision makers (National Crisis Coordination Centre, ministerial crisis centres, the Safety Regions, etc). Unfortunately, crisis organisations operate in just those circumstances in which failure of telecommunication services is most likely. A breach of a dike or an explosion in a chemical installation increases the odds that the supply of electricity or telecom switches will fail.

It is therefore important that organisations in general, and crisis organisations in particular, understand the vulnerabilities and dependencies of the telecom services they use. This document describes a method, called Raster, to assist in this understanding.

The goal of Raster is that the organisation becomes less vulnerable to telecom failures. To reduce the vulnerability, the organisation must first understand what can go wrong with each telecom service they use. Also, these risks must be ranked, so that the most pressing risks can be addressed first. Raster helps a team of analysts to map and investigate one or more telecom services for an organisation. The result is a report, showing which risks should be addressed first, and why. Selection and execution of countermeasures is the next logical step, but is not part of the Raster method.

1.2 The Raster method

Incidents with availability of telecom services often happen because of component failures: an underground cable is damaged by a contractor, a power failure causes equipment to shut down. To prepare for these incidents, the organisation must first realise that the cable and equipment exist. An important part of the Raster method is therefore to draw a diagram showing all components involved in delivering the service.

Incidents can also happen when a single event leads to the simultaneous failure of two or more components. For example, two cables in the same duct can be cut in the same incident, or a software update can cause several servers to misbehave. These failures are called *common cause failures*, and they are dangerous because their impact can be quite large.

Major steps in the Raster method are to draw service diagrams, and to assess the likelihood and potential impact of single and common cause failures. However, unlike other methods Raster does not take a narrow numerical approach to assessing risks.

Risks with low probability and high effects are especially important. These rare but catastrophic events have been called “black swans”. Raster helps to uncover black swans in telecom services.

Risk assessments are always in part subjective, and information is hardly ever as complete as analysts would like it to be. This does not mean that biases and prejudices are acceptable. Raster tries to nudge analysts into a critical mode of thinking. Uncertainty is normal, and assessments can be explicitly marked as “Unknown” or “Ambiguous” if a more specific assessment cannot be made. Raster can be applied even when much of the desired information on the composition of telecom networks is unavailable or unknown. Missing information can be gradually added.

To avoid a narrow risk assessment, the Raster method is applied by a team of experts, each having his own area of expertise. Raster facilitates cooperation between experts of different backgrounds.

Raster facilitates the construction of a recommendation using a tested methodical analysis. This recommendation is not just based on the technical aspects of failure of telecoms services, but also takes account of the societal impact of failures, and of risk perceptions of external stakeholders.

The following parties are involved in applying the Raster method.

- The crisis organisation: the method is executed on request of a crisis organisation. This organisation is the requesting client of the study.
- The analysts: the method is executed by a group of professionals. It is essential that this group consists of multiple people. Not only does a single person seldom possess all required knowledge, it is also important that the study leads to an

objective and impartial assessment, as much as possible free from personal preferences or personal blind spots.

The team needs to encompass knowledge on crisis management, crisis communications, and technical aspects of telecommunication networks and services. Additionally, it will be useful if team members have some experience with risk assessment, and with the Raster method in particular. Because of this range of knowledge it will typically be useful to include representatives of the crisis organisation in the team of analysts.

- The sponsor: the person or entity representing the crisis organisation for the purpose of the study. Typically this will be an official of the crisis organisation.
- The decision makers: the output of the method is a set of recommendations and supporting argumentation that serve as the basis for the selection of risk treatment decisions. Responsibility for the selection does not belong to the analysts, but to the decision makers. The decision maker can be sponsor, but these roles can also be separate.
- The external stakeholders: this category includes all parties that are not part of the crisis organisation and not involved in the use of telecom services, but do have interests that may be harmed by the risks or chosen risk treatments. External stakeholders may be ‘the public’ in general, or a specific group such as those people living in the neighbourhood of a facility.

1.3 About this manual

This manual is for the professionals who will execute the Raster method. It explains the method and provides guidance. These professionals can either be telecom experts or experts in any other field whose expertise is needed.

In this manual, the words ‘must’, ‘should’, and ‘may’ have a well-defined meaning.

- *Must* indicates a compulsory aspect of the Raster method; under no circumstances can the activity be omitted.
- *Should* indicates a recommended activity, that should only be omitted if the implications are fully understood. This must be a conscious decision.
- *May* indicates a suggested but optional activity, that can be included or omitted at will.

Examples, notes and tips are typeset in text boxes.

This would be an example, note, tip or shortcut.

1.4 Outline of this manual

Chapters 3 to 6 describe the Raster method; chapters 8 to 13 describe the Raster tool that aids the creation of diagrams and the analysis of Single Failures and

Common Cause Failures. When executing an analysis using Raster, you will proceed as in the figure below.

2 The Raster method

General outline of the Raster method and telecom service diagrams.

2.1 Outline

When using the Raster method, you and the rest of your team will perform a number of tasks. The method will guide you through these tasks in a methodical way, and the Raster tool will assist you in recording your progress. Based on your collective knowledge and expert judgement you will make estimates about the likelihood and impact of various vulnerabilities affecting the telecom services. Based on this analysis, you and your team will draft suitable risk treatment recommendations. The result of your efforts is a report that can be used by a decision maker to take informed business decisions about accepting, reducing, or avoiding the risks.

Raster consists of four stages, shown in the figure below.

1. Initiation and preparation
2. Single failures analysis
3. Common cause failures analysis
4. Evaluation

1. The Initiation and Preparation stage describes the scope and purpose of the assessment. Which telecom services are involved, which users can be identified, who are external stakeholders, and what are the characteristics of the environment in which these services are used?
2. The Single Failures Analysis stage creates a telecom service diagram for each telecom service in use. These diagrams describe the most relevant telecommunication components, including cables wireless links, and equipment items. These components are potentially vulnerable. The diagram does not have to be complete in all details. Parts of networks that are less relevant can be captured using a single “cloud” (unknown link). For all components an assessment of all applicable vulnerabilities is done. Only independent, single failures are taken into account during this stage.
3. The Common Cause Failures Analysis stage takes closer look at failure causes that lead to the failure of multiple components at once. One example is that of independent telecom services that both have a cable in the same underground duct. A single trenching incident may cut both cables at the same time, causing both services to fail. Another example is a large-scale power outage, causing equipment over a large area to fail simultaneously.
4. The Risk Evaluation stage contains the risk evaluation and creation of the final report. The overall risk level is assessed, and recommendations are done for risk treatment. These recommendations take into account the possible reactions of external stakeholders. The recommendations and their supporting argumentation form the final output of the Raster method.

Chapters 3 to 6 describe each stage in detail.

To facilitate the creating of diagrams and analysis of single and common cause failures, the Raster tool is available. This tool is described in the second part of this document, starting from Chapter 8. In principle, stages 2 and 3 can be used without the tool. However, the tool comes highly recommended and this manual assumes that the method will be used together with the tool.

2.2 Telecommunication service diagrams

Diagrams are central to the Raster method. A telecom service diagram describes the physical connectivity between components of a telecom service. Diagrams consist of nodes that are connected by lines. Each line represents a direct physical relation. It indicates that the nodes are attached to each other. There cannot be more than one line between two nodes; nodes are either connected or they are not.

Lines are not the same as cables. When two equipment items are connected via a cable, three nodes are used as in the picture above. The line between equipment

and cable shows a physical connection: the cable is plugged into the equipment.

There are five types of nodes, each identified by its unique shape.

2.2.1 Actors

Actors represent the (direct) users of telecom services. An actor can represent a single individual, or a group of individuals having the same role, e.g. 'journalists' or 'citizens'. Maintenance personnel are not modelled as actors, as they do not participate in communication.

An actor can only be connected to components of type 'equipment' or 'unknown link'. Actors cannot be connected directly to wired or wireless links, and the Raster tool will not allow such connections.

There must be at least two actors in the diagram. There must at least be a person communicating, and one other person to communicate with.

2.2.2 Wired links

Wired links represent passive, physical cables, including their connectors, fittings and joints but excluding any active components such as amplifiers or switches. Fiber optic cables, coaxial cables, and traditional telephony copper pairs are typical examples of wired links. The two equipment items connected by the link are not part of the wired link itself, and need to be included in the model separately, either as equipment items or unknown links.

wired link

Each wired link has exactly two connections, each to a component of type either 'equipment' or 'unknown link'. To connect a wired link to an actor, wireless link, or an other wired link, place an equipment node in between.

Each wired link has some fixed capacity, a physical location (including a height above or below ground level). These properties need to be known in sufficient detail.

2.2.3 Wireless links.

Wireless links represent direct radio connections, excluding any intermediate components. The transmission and reception installations are not part of the wireless link, and have to be modelled separately as equipment items. A wireless link can connect two or more nodes.

Each wireless link has a fixed capacity, but unlike wired links a wireless link does not always have a fixed location. Transmitters and receivers can be mobile or nomadic. The coverage area depends on factors such as transmission power and antenna properties. Wireless links have a fixed frequency or band. All of these properties need to be described in sufficient detail.

Each wireless link has at least two connections, each to a component of type either 'equipment' or 'unknown link'. It can have more than one, as in the example above. To connect a wireless link to an actor, equipment, or an other wireless link, place an equipment node in between.

2.2.4 Unknown links

Unknown links (cloud shapes) represent parts of networks for which insufficient information is available, or that do not need to be described in detail. Unlike wired and wireless links, that represent a single communication channel, unknown links are composed of equipment and wired and wireless links.

Because unknown links are collections of equipment and wired and wireless links, they can be used in any place where these nodes can be used. In short, unknown links can connect to any other node type. Also, unknown links can be connected to any number of nodes.

2.2.5 Equipment

Equipment nodes represent all other physical components of telecom networks, such as switches, exchanges, routers, amplifiers, radio transmitters, radio receivers etc. An equipment node may model a single piece of equipment or an entire installation.

equipment

Each equipment node must be connected with at least one other component. An equipment node cannot be connected directly to another node of type 'equipment'.

2.2.6 Example

The figure below shows an example of a valid telecom service diagram. The diagram shows three actors, communicating via telephony. Two actors are connected to the same private exchange (PABX); the third actor is abroad. One actor uses a wireless DECT handset and base station, the others use fixed handsets. We have no knowledge (yet) of the other portions of the network, other than that some PABX must exist, and some kind of international telephony network to facilitate the calls.

3 Stage 1 — Initiation and preparation

Define shared purpose and bounds to the study.

Before the study is started its scope must be made clear to the analysts and to the sponsor. The responsibilities and tasks of the crisis organisation must be described in some detail. Also, the position of this crisis organisation within the wider system of crisis response must be laid out.

In stage 1 you will collect the information that you need to complete the other stages. The result is a report and agreement from the sponsor to proceed.

The Initiation and Preparation stage consists of the following steps:

1. Identify telecom services
2. Identify actors
3. Describe disaster scenarios
4. Create Stage 1 report
5. Obtain approval from sponsor

3.1 Identify telecom services

Create a list of all telecommunication services that are used by the crisis organisation. This list must be exhaustive. If a service is accidentally omitted, no risk assessment will be performed on it, and dependencies between the service and other services will not be discovered. As a result, decision makers may take unnecessary or ineffective countermeasures, or overlook necessary countermeasures.

To create the list of telecom services, the following information sources may be useful:

- The initiating problem statement, project initiation document, or request for proposals.
- Interviews with executives and operational staff from the crisis organisations.
- Observation of operational staff in exercises or real-life operations.
- Disaster preparedness plans.
- Reports or evaluations of past exercises.
- Internal formal procedures, operational guides, process manuals.
- Reference materials used during crisis response.

Briefly describe each telecom service. At this stage it is not yet necessary to describe the technical implementation, but if information is available on such items as handsets, terminals, or links, then this should be included in the descriptions.

If a telecom service acts as backup to some other telecom service, or when the service itself has fallback options, then these must be described as well.

The descriptions must also include the relevance of the telecom service to the operations of the crisis organisation. That is, is the service essential, or merely a ‘nice to have’?

It will also be useful at this stage to start a glossary of abbreviations and definitions of special terms that may not be clear to all analysts, or to the sponsor.

3.2 Identify actors and external stakeholders

List, for each telecom service, the actors who may make use of that service. Main actors are members of the crisis organisation. All other actors are secondary actors. Actors can be the initiating party of communication session (calling party) or the receiving party (called party), or both.

List all external stakeholders to the crisis organisation.

Actors and external stakeholders may be identified using the same information sources as listed above for telecom services.

3.3 Describe disaster scenarios

Before the analysis can start, it must be clear to which threats this crisis organisation may be exposed. For example, the in-company fire service in charge of chemical plant safety will be confronted with different potential disasters than a crisis team controlling the spread of agricultural diseases. The latter is unlikely to be affected by violent destruction of hardware. Consequently, the threats to their telecom services will be very different in nature.

The threats to telecom services and their mechanisms must be described in as much detail as possible. Disaster scenarios describe the threats, their effects and mechanisms, their likelihood, and the required response from the crisis organisation.

In the Netherlands tornados seldom lead to damage to infrastructures. Typically, the threat of tornados will therefore be excluded from disaster scenarios. Flooding from sea or riverbeds, however, are quite common, and will likely be included.

For some studies intentional human-made events (crime, terrorism) are highly relevant. For other studies it may suffice to focus on accidental events only. The scope of the study need not be limited to technical aspects.

When describing a disaster, the effects that it will have on telecom components is the most important part. To better understand the reactions of the general public it may be useful to also include some graphic descriptions of events that could be experienced by citizens, or that could be published in the media. This may facilitate the assessment of social risk factors in the Risk Evaluation stage.

It may be possible to reuse disaster scenarios from previous risk assessments, thus shortening the amount of work needed.

3.4 Create stage 1 report

The results from Stage 1 must be recorded because the analyst will need to refer to this information during subsequent stages.

The following is a common outline of the output document of the Initiation and Preparation stage. This report forms the introduction to the final report (see section 6.4).

1. Executive summary to the Stage 1 report.
2. About the crisis organisation (internal scope):
 - a. Position within wider system of crisis response.
 - b. Sponsor, decision makers, and analysts.
 - c. Roles, tasks, and responsibilities of the crisis organisation.
 - d. Telecom services used, with a description of the implementation, role and purpose during crisis response, and fallback and backup options.
 - e. Actors, including main actors, and their roles, tasks, and responsibilities.
3. About the environment of the crisis organisation (external scope):
 - a. Disaster scenarios, with descriptions.
 - b. External parties with whom the main actors may communicate, and other external stakeholders.
4. Glossary

3.5 Obtain approval from sponsor

All analysts must participate in a review of the Stage 1 report. All analysts must agree on its contents by consensus.

The Stage 1 report must then be presented to and discussed with the sponsor. The list of telecom services may contain unexpected services. The unexpected appearance of a service is informative, since it indicates that the risk assessment and preparation of the crisis organisation are insufficient, and that disaster response plans are incomplete.

The results of the Initiation and Preparation stage determine to a large extent the course of the risk assessment in the later stages. It is therefore important that the sponsor also agrees to the outcome of this stage, and gives formal agreement to the resulting documentation. As a consequence, the documents must be understandable to non-experts. A glossary may be helpful to that effect. Also, an executive summary should be written.

4 Stage 2 — Single failures analysis

Describe telecom service networks and analyse vulnerabilities of components.

In this stage you will create a telecom service diagram for each telecom service, and assess the vulnerabilities on each of its components. This will give you a good understanding of the inner workings of each telecom service, and a first impression of its risks.

The result will be recorded in the Raster tool: telecom service diagrams and assessment of Frequency and Impact on vulnerabilities of diagram components.

The Single Failures Analysis stage consists of the following steps:

1. Update the checklists of vulnerabilities
2. Draw initial diagrams
3. Analyse the vulnerabilities of components (assess frequency and impact)
4. Expand unknown links
5. Review

4.1 Update the checklists of vulnerabilities

Based on the disaster scenarios that were described in Stage 1, you must describe the most common vulnerabilities of network components. Checklists are used for this. A checklist contains the name and description of the most common vulnerabilities. Good checklists make the analysis process faster and easier.

Create a fresh Raster project (see section 8.3.1), and inspect the predefined checklist for each type (see section 9.2). Add new vulnerabilities as deemed necessary. Include vulnerabilities that apply to most components of that type; omit vulnerabilities that only apply to a few components. The checklists do not have to be complete; any particular network component may have specific vulnerabilities that do not occur in the checklist. However, when the most common vulnerabilities are included in checklists, few special cases need to be considered.

There are three checklists, one each for equipment, wired and wireless links. For actor components no checklist exists. Vulnerabilities of actors are outside the scope of the Raster method. Also, unknown links do not have a separate checklist. They may contain any of the other component types, and therefore all vulnerabilities of the three checklists may apply to unknown links.

Vulnerabilities of actors are not taken into account. For example, Raster does not handle an actor misinterpreting a received message. However, configuration errors, incorrect handling of handsets or cyber crimes can be taken into account. These vulnerabilities are modelled in Raster as part of equipment components, not as part of the actor responsible for them. Maintenance personnel are not included in the diagrams as actors.

4.2 Draw initial diagrams

In the Raster tool, create a diagram tab for each telecom service (see Section 9.3.1).

Then, for each telecom service, draw an initial diagram based on the information that is currently available. The diagrams will likely not be very detailed yet. At the very least all actors involved with the service must be drawn. Note that it is always possible to create a diagram; if absolutely no information is available beyond the actors involved then the actors can simply be connected using an unknown link (“cloud” symbol). Drawing and editing diagrams using the Raster tool is explained in Section 9.3.

When creating diagrams, the following guidelines may be helpful:

- A cable containing multiple fibers or strands should be modelled as a single wired link. Two cables in the same duct should be modelled by two wired links in the diagram.
- Point-to-multipoint connections should be modelled using a single wireless link, but may sometimes be more conveniently modelled using separate wireless links to each receiving node. If you know in advance that the link to each individual node is subject to identical risks, then for simplicity a single wireless link should be used.
- Equipment components can be a single device, or an entire installation. For example, a small telephone exchange may be modelled as a single equipment node. However, installations such as these contain multiple cables and sub-components. Often it is not necessary to model these cables and equipment items separately. When an installation is separated over multiple rooms or when wireless links are used then the sub-components should be modelled separately. Alternatively, an unknown link may be used instead of an equipment item.

4.3 Analyse the vulnerabilities of components

This activity must be performed for each component in turn. Each step, a component is selected for analysis.

4.3.1 Add and remove vulnerabilities

Inspect the listed vulnerabilities of the component. The initial list is a copy of the generic checklist for that type. Other vulnerabilities may exist that were not in the checklist. These vulnerabilities must be added. The disaster scenarios that were prepared in Stage 1 must be used as guidance in decisions to add vulnerabilities.

Example: Telecommunication satellites are vulnerable to space debris. This vulnerability does not apply to any other kind of equipment, and will therefore not be in the equipment checklist. On the other hand, satellites are not vulnerable to flooding. Therefore “Collision with space debris” must be added, and “Flooding” must be removed from the list of satellite vulnerabilities.

A vulnerability must not be removed unless it is clearly nonsensical, e.g. configuration errors on devices that do not allow for any kind of configuration, or flood damage to a space satellite. To be removed, a vulnerability must be physically impossible, not just very unlikely in practice. In all other cases the frequency and impact of the vulnerability should be assessed (although they can both be set to Extremely low), and the vulnerability must be part of the review at the end of Stage 2.

It is important that vulnerabilities that are merely unlikely but not physically impossible are retained in the analysis, because such vulnerabilities could have an extremely high impact. Low-probability/high-impact events must not be excluded from the risk analysis.

4.3.2 Assess vulnerabilities

When the list of vulnerabilities for the component is complete, each vulnerability must be assessed. The analysts, based on their collective knowledge, estimate two factors:

1. the likelihood that the vulnerability will lead to an incident (its frequency), and
2. the impact of that incident.

Both factors Frequency and Impact are split into eight classes, summarised in Tables 4.1 and 4.2. The classes do not correspond to ranges (a highest and lowest permissible value); instead they mention a typical, characteristic value for the class. The selection of the proper class may require a discussion between analysts. Analysts must provide convincing arguments for their choice of class.

Sometimes a factor (a likelihood or impact) is extremely large, or extremely small. Extremely large values are not simply very big, but too big to fit in the normal scale, unacceptably high and intolerably high. Likewise, extremely small values are outside the scale of normal values, and sometimes may safely be ignored. Extreme values fall outside the normal experience of analysts or other stakeholders, and normal paths of reasoning cannot be applied.

If no consensus can be reached between the analysts, the class *Ambiguous* must be assigned. In the remarks the analysts should briefly explain the cause for disagreement, and the classes that different analysts would prefer to see.

A limited amount of uncertainty is unavoidable, and is normal for risk assessments. However, when uncertainty becomes too large, so that multiple classes could be assigned to a factor the class *Unknown* must be assigned.

The Raster tool assists in recording the analysis results. The tool will also automatically compute the combined vulnerability score for each vulnerability, and the overall vulnerability level for each node (see sections 9.3.11 and 10, and section 13.5 for technical details).

Do not blindly trust your initial estimate of frequency and impact. You must not rely only on information that confirms your estimate, but also actively search for contradicting evidence.

4.3.3 Assess frequency

The factor Frequency indicates the likelihood that the vulnerability will lead to an incident with an effect on the telecom service. All eight classes can be used for Frequency (see Table 4.1).

A frequency of “once in 50 years” is an average, and does not mean that each 50 years an incident is guaranteed to occur. It may be interpreted as:

- The average timespan between incidents on a single component is 50 years.
- For a set of 50 identical components, each year on average one of them will experience an incident.
- Each year, the component has a 1 in 50 chance of experiencing an incident.

When the life time of a component is 5 years (or when the component is replaced every 5 years) the frequency of a vulnerability can still be “once in 500 years”.

Example: a component is always replaced after one year, even if it is still functioning. On average, 10% of components fail before their full year is up. The general frequency for this failure is therefore estimated as “once in 10 years” even though no component will be in use that long.

Note that this value is between the characteristic values for High and Medium. The analysts must together decide which of these two classes is assigned.

Use the following three-step procedure to determine the factor Frequency:

1. Find the frequency class that applies to this type of node in general.
This can be based on, for example, past experience or expert opinion. If available, MTBF (mean time between failures) figures or failure rates should be used.
2. Think of reasons why this particular node should have a lower or higher frequency than usual.

Existing countermeasures may make the frequency lower than usual. For example, if an organisation already has a stand-by generator that kicks in when power fails, then the frequency of power failure incidents is thereby reduced. Remember that the frequency does not reflect the likelihood that the vulnerability is triggered, but the likelihood that the vulnerability will lead to an incident.

For some components monitoring can detect failures that are imminent before they occur. This also will reduce the frequency of incidents. Another example is the use of premium quality components, or secure and controlled equipment rooms. All of these measures make incidents less likely.

<i>Class</i>	<i>Value</i>	<i>Symbol</i>
High	Once in 5 years.	H
	For 100 identical components, each month 1 or 2 will experience an incident.	
Medium	Once in 50 years.	M
	For 100 identical components, each year 2 will experience an incident.	
Low	Once in 500 years.	L
	For 100 identical components, one incident will occur every five years.	
Extremely high	<i>Routine event. Very often.</i>	V
Extremely low	Very rare, but not physically impossible.	U
Ambiguous	Indicates lack of consensus between analysts.	A
Unknown	Indicates lack of knowledge or data.	X
Not yet analysed	Default. Indicates that no assessment has been done yet.	–

Table 4.1: Characteristic values for frequency classes.

The disaster scenarios may be an indication that the frequency should be higher than usual. In crisis situations it is often more likely that an incident will occur. For example, power outages are not very common, but are far more likely during flooding disasters. These disasters themselves are very uncommon. The overall frequency is therefore determined by:

- the likelihood of power outages during normal circumstances, and
- the likelihood of power outages during a flood, combined with the likelihood of flooding.

3. Decide on the frequency class for this particular node.

Typically either Low, Medium, or High will be used. If neither of these accurately reflect the frequency, one of the extreme classes should be used. If no class can be assigned by consensus, one of Ambiguous or Unknown should be used.

4.3.4 Assess impact

The factor Impact indicates the severity of the effect when a vulnerability does lead to an incident. This severity is the effect to the service as a whole, not its effect to the component that experienced the vulnerability. For example, a power failure will cause equipment to stop functioning temporarily. This is normal, and in itself of little relevance, unless it has an effect on the availability of the telecom service. The power failure could cause the service to fail (if the equipment is essential), but could also have a no effect at all (if the equipment has a backup). Or any effect in between.

Only the effects on the telecom service must be taken into account in this stage. Loss of business, penalties, and other damage are not considered, but may be relevant during risk evaluation (see Section 6.3.2).

The damage may be caused by an incident that also affects other components of the same telecom service. For example, a cable may be damaged by an earthquake; the same earthquake will likely cause damage to other components as well. However, this additional damage must not be taken into account. Only the damage resulting from the damage to this component must be considered. The next stage, common cause failures analysis, takes care of multiple failures due to a single incident.

The impact of some vulnerability on a component covers:

- only effects to the service, not the effects to the component itself,
- only effects to the service, not subsequent damage to the organisation,
- only effects due to damage this single component, not effects due to the failure scenario.

All eight classes can be used for Impact. Characteristic values for the classes high, medium, and low are given in Table 4.2.

Use the following three-step procedure to determine the factor Impact:

1. Choose the impact class that most accurately seems to describe the impact of the incident.
2. Think of reasons why the impact would be higher or lower than this initial assessment.

Existing redundancy can reduce or even annul the impact. For example, a telecom service may have been designed such that when a wireless link fails, a backup wired link is used automatically. The impact of the wireless link failing is thereby reduced.

Monitoring and automatic alarms may reduce the impact of incidents. When incidents are detected quickly, repairs can be initiated faster. Keeping stock of spare parts, well trained repair teams, and conducting regular drills and exercises all help in reducing the impact of failures and must be considered in the assessment. On the other hand, absence of these measures may increase the impact of the incident.

3. Decide on the impact class.

Typically either Low, Medium, or High will be used. If neither of these accurately reflect the impact, one of the extreme classes should be used. If no class can be assigned by consensus, one of Ambiguous or Unknown should be used.

It typically does not matter for the selection of impact class whether some or all actors are affected. All actors are important; they would not appear in the diagram

otherwise. However, if the analysts agree that only very few actors are affected they can select the next lower class (e.g. Low instead of Medium).

The meaning of “short-term” and “long-term” depends on the tasks and use-cases of the actors. A two-minute outage is short-term for fixed telephony but long-term for real-time remote control of drones and robots.

“Degradation” means that actors notice reduced performance (e.g. noise during telephone calls, unusual delay in delivery of email messages), but not so much that their tasks or responsibilities are affected.

<i>Class</i>	<i>Value</i>	<i>Symbol</i>
High	Partial unavailability, if unrepairable. Total unavailability, if long-term.	H
Medium	Partial unavailability, if repairable (short-term or long-term). Total unavailability, if short-term.	M
Low	Noticeable degradation, repairable (short-term or long-term) or unrepairable.	L
Extremely high	Very long-term or unrepairable unavailability of the service.	V
Extremely low	Unnoticeable effects, or no actors affected.	U
Ambiguous	Indicates lack of consensus between analysts.	A
Unknown	Indicates lack of knowledge or data.	X
Not yet analysed	Default. Indicates that no assessment has been done yet.	–

Table 4.2: Characteristic values for impact classes.

“Partial unavailability” means severe degradation or unavailability of some aspects of the service, such that actors cannot effectively perform some of their tasks or responsibilities. For example: email can only be sent within the organisation; noise makes telephone calls almost unintelligible; mobile data is unavailable but mobile calls and SMS are not affected. Actors can still perform some of their tasks, but other tasks are impossible or require additional effort.

“Total unavailability” means that actors effectively cannot perform any of their tasks and responsibilities using the telecom service (e.g. phone calls can be made but are completely unintelligible because of extremely poor quality).

“Extremely high” means that if the incident happens the damage will be so large that major redesign of the telecom service is necessary, or the service has to be terminated and replaced with an alternative because repairs are unrealistic.

4.3.5 Assessing all vulnerabilities on a component

The overall vulnerability level of a component is defined as the worst vulnerability for that component. If some of the vulnerabilities are not assessed (no frequency or impact have been set on them), they will not contribute to the overall vulnerability level. It can thus be a useful time-saver to skip assessment of unimportant vulnerabilities.

It is very important that all vulnerabilities with High and Extremely high impact are assessed fully. This is true even when their Frequency is low.

4.4 Expand unknown links

When an unknown link receives an overall vulnerability level of Ambiguous or Unknown, the analysts must decide whether or not to expand the node. Expansion means that the internal make-up of the node is examined; the unknown link is removed from the diagram, and its constituent parts are added to the diagram as individual equipment items, wired and wireless links, and possibly further unknown links. Expansion adds more detail to the model, and results in additional diagram components. The vulnerabilities to these new components must also be analysed, as for any other diagram component.

It is not always necessary to expand unknown links. If the analysts think that the effort involved in expansion is too large, or that it will not lead to more accurate or insightful results then expansion should be omitted.

4.5 Review

When all components have been analysed, a review must take place. All analysts must participate in this review. The purpose of the review is to detect mistakes and inconsistencies, and to decide whether the Single Failures Analysis stage can be concluded.

If any of the components has an overall vulnerability level of Ambiguous or Unknown, the analysts must decide whether or not to conduct further investigation, in order to assess the vulnerabilities to that node with greater certainty. If the analysts think that the effort involved is too large, or that it will not lead to more accurate or insightful results then the component should be left as is.

If the analysts decide to redo some part of the Single Failures Analysis stage, then they should again perform a review afterwards. This review may be omitted when the analysts agree that all changes are minor.

5 Stage 3 — Common cause failures analysis

Determine and analyse common cause failures.

A common cause failure is an event that leads to the simultaneous failure of two or more components. For example: two cables in the same duct can both be cut in a single incident; multiple equipment items may be destroyed in a single fire.

For a common cause failure to happen, the affected components must be within range of each other, according to a critical property. For physical failure events such as fire and flooding, this property is geographical proximity: the components must be sufficiently close to be affected simultaneously. For configuration mistakes it is the similarity in maintenance procedures. For software bugs it is whether related firmware versions are used, regardless of geographical distance. Other events may have different critical properties.

For each failure scenario, the critical property has a maximum effect distance. Two equipment items can only be affected by a minor fire when they are in the same room; for a major fire the effect distance is larger, but still limited to perhaps a single building. Flooding has a much larger effect area, and two components must be further apart to be immune from flooding as their common failure cause.

In stage 3 you will make groups of components that fall within the same range of a critical property. You will do this for each vulnerability separately. For each cluster you will then assess the Frequency and Impact of a common cause failure affecting the components in that cluster. The clusters and their assessments will be recorded in the Raster tool. The result is an improved and refined risk assessment.

The Common Cause Failures Analysis stage consists of the following steps:

1. Create clusters
2. Analyse each cluster
3. Expand unknown links
4. Review

5.1 Create clusters

The Raster tool automatically lists each vulnerability in use, provided that that vulnerability occurs for at least two components. For each such vulnerability, the analyst must create clusters based on the critical property.

Example: clusters based on *geographical proximity* can be used for fire, flood, power outage, cable breaks, and radio jamming (per frequency band).
Clusters based on *organisational boundaries* can be used for equipment configuration, ageing, and software bugs.

Initially, the Raster tool places all components that have the same vulnerability in a single cluster. Based on the effect distance of failure scenarios further subdivisions can be made, such that:

- each cluster represents a class of failure scenarios that are similar in location and effect area.
- a failure scenario for a cluster can never affect components outside that cluster.
- any two components in the same cluster may be affected by the same failure scenario simultaneously.

It is possible for a larger cluster to entirely include a smaller cluster. Clusters may thus be nested. All nodes in a subcluster are members of their parent cluster as well.

For example, the figure to the right shows an office floor plan with two equipment rooms. Three possible clusters are:

1. Equipment room 1 – small fires, affecting components in equipment room 1 only.
2. Equipment room 2 – small fires, affecting components in equipment room 2 only.
3. Entire office – large fires, affecting all components in all rooms.

Cluster 3 then contains subclusters 1 and 2. Note how each cluster is specific to one vulnerability (fire), and covers scenarios that have the same location and effect area.

Chapter 11 describes how the Raster tool can be used to create clusters.

5.2 Analyse each cluster

To analyse a cluster the two factors Frequency and Impact must be assessed. This is done in the same way as for single failures (see section 4.3).

In this stage, the factor Frequency reflects the likelihood that *two or more* components in that cluster are affected by the same threat event. The factor Impact still indicates the overall effect on the telecom service, when the threat event leads to an incident.

The Raster tool will automatically compute the vulnerability level of any parent clusters, including the top level vulnerability.

5.3 Expand unknown links

When a cluster containing unknown links receives an overall vulnerability level of Ambiguous or Unknown, the analysts must decide whether or not to expand those

unknown links. This is analogous to expansion in the Single Failures Analysis stage (see section 4.4).

Note that it is not always necessary to expand unknown links. If the analysts think that the effort involved in expansion is too large, or that it will not lead to more accurate or insightful results then expansion should be omitted.

Expansion adds new components to the diagram. These new components need to be analysed for single failures. This means that part of Stage 2 needs to be redone for these components. It also means that some clusters receive new member nodes. The analysis of these clusters must be revisited.

5.4 Review

During the final review all analysts must discuss the results of the analysis of single and common cause failures. Special care must be taken to ensure that all assessments are consistent. The next stage must only be started when all analysts agree on the analysis results.

If any of the clusters has an overall vulnerability level of Ambiguous or Unknown, the analysts must decide whether or not to conduct further investigation, in order to be able to assess the common cause failures within that cluster with greater certainty. If the analysts think that the effort involved is too large, or that it will not lead to more accurate or insightful results then the component should be left as is.

If the analysts decide to redo major parts of the common cause failures analysis, then they should perform another review afterwards.

6 Stage 4 — Risk evaluation

Prioritise and evaluate risks, and make treatment recommendations.

When all single and common cause failures have been analysed, a list of the most serious risks can be made. The Raster tool assists the initial effort for this stage. Quick wins can be determined automatically, and simple “what if” analysis is available.

During this stage you make a judgement of which risks you consider to be too large. You write down arguments for your choice and propose risk treatments. You take into consideration both your assessment of vulnerabilities that affect the availability of telecom services and your expectation of reactions of other stakeholders to risk treatments. The result is a report to the sponsor, outlining, explaining and justifying your recommendations.

The Risk Evaluation stage consists of the following steps:

1. Determine longlist
2. Reduce longlist to shortlist
3. Make treatment recommendations, considering social risk factors
4. Prepare final report

6.1 Determine longlist

Based on the information presented by the Raster tool (see section 12), a longlist of the most serious risks must be compiled. These risks are:

- the combination of a single vulnerability and a single component, such as “power failure at the PABX”, or
- the combination of a single common cause failure vulnerability and a single cluster, such as “fire at equipment in the facilities room”.

It is up to the analysts to judge which risks are serious enough to be placed on the longlist. However, the list should include the “quick wins” reported by the Raster tool (see section 12.1). Quick wins are those vulnerabilities that by themselves determine the overall vulnerability level of a component. Reducing that vulnerability would immediately reduce the overall level.

Other good candidates for inclusion on the longlist are those risks that were computed as Extremely high or High, as well the risks that were computed as Ambiguous or Unknown.

6.2 Reduce longlist to shortlist

The longlist must be prioritised. Prioritisation requires more information than can be found in the diagrams and vulnerability assessments. For example, information on control relationships between components, or information about redundancy,

cannot be found in the diagrams but is very important for risk prioritisation. Also, telecom services are not all equally important. Therefore, a risk that was assessed as “high” occurring in a service that is useful but non-essential may be listed below a risk that was assessed as “medium” to a vital service. The priority may further be affected by the service acting as backup to another service, or having fallback options itself.

All analysts must collectively examine each risk on the longlist. Based on consensus, risks may be raised or lowered on the list, or may be removed altogether. The result of this process is a prioritised shortlist of risks for which the analysts agree that risk treatment is warranted.

6.3 Make treatment recommendations

It is not the responsibility of the analysts to decide on how risks on the shortlist will be treated. The sponsor or decision maker will be responsible for these measures. However, the analysts do have the responsibility for providing them guidance, and to make reservations for the uncertainty in their assessments and limits to their knowledge.

For each risk on the shortlist, the analysts must give risk treatment recommendations. It is impossible to give a procedure for this, as the suitable treatment for a risk depends very much on the type of service, the nature of the risk, and the circumstances of the crisis organisation.

6.3.1 Select risk treatment option

In general, four general risk treatment options exist:

1. **Avoid.** Remove the risk completely (proaction), or discontinue the use of the component or service altogether. Proaction means eliminating structural causes of accidents to prevent them from happening in the first place (e.g. avoiding radio interference by replacing a wireless link by a wired link). When discontinuing an entire service, an alternative service will often be available. However, you should be careful to replace a service with known risks for a new one with unknown risks.

Even when no alternative is available it may still be worth considering discontinuing use of the telecom service when the risk cannot be avoided. Rather than using a service that may fail unexpectedly, it may be preferable to not use the service at all to avoid unpleasant surprises at inopportune moments during crisis response.

2. **Reduce.** Make the risk more acceptable, by reducing either its likelihood (frequency) or impact. These activities encompass prevention and preparation. Prevention means taking measures beforehand that aim to make accidents less likely, and to limit the consequences in case incidents do occur (e.g. by imposing smoking restrictions and using fire-retardant materials). Preparation means

ensuring the capacity to deal with accidents and disasters in case they do happen (e.g. by holding regular fire drills).

3. **Transfer.** Pass the risk to another party. Typical examples of risk transfer are insurance, or maintenance contracts whereby faulty equipment is replaced with spares on short notice. Risk transfer in effect buys certainty, by transferring the uncertainty to another party in return for payment.
4. **Retain.** Accepting the risk, in an informed decision. Reasons for accepting risks may be that other options would be too costly, that the likelihood is deemed to be very low, or simply the lack of suitable alternatives. In all cases it is much preferable to knowingly accept a risk rather than being confronted with it.

<i>Factor</i>	<i>Description</i>
Artificiality, immorality	“Unnaturalness” of risk sources.
Benefits	Tangible and intangible beneficial effects.
Blame	Responsibility for damages attributable to some actor.
Catastrophic potential	Fear of sudden, disruptive, large effects.
Children	Amount of risk exposure faced by children in general.
Familiarity	Extent to which the risk is perceived as common and well known.
Fear	Characterises the amount of fear.
Institutional control	Close, effective monitoring of risks by authorities, with the option of intervention when necessary.
Media exposure	Amount of attention by (social) media.
Mobilisation	Potential for protests and active opposition.
Personal control	Level of control that an individual stakeholder can exercise.
Violation of equity	Discrepancy between those who enjoy the benefits and those who bear the risks.
Voluntariness	Amount of free choice an individual has in being exposed to the risk.

Table 6.1: List and description of social risk factors.

6.3.2 Assess social risk factors

The draft treatment of risks on the shortlist may lead to criticism by other stakeholders. The opinions of these stakeholders must be considered before final treatment recommendations are formulated. Otherwise, decision makers may unexpectedly have to deal with societal opposition, possibly forcing them to opt

for a sub-optimal treatment that is nevertheless more acceptable to external stakeholders. Analysts must therefore assess additional risk factors that influence risk perception and risk acceptance by third parties. See Table 6.1.

Artificiality applies to situations where people oppose a technology because it is unnatural. For example, electromagnetic radiation from mobile telephony base stations is more often considered ‘harmful’, whereas natural sunlight is more often considered ‘healthy’. However, there is scientific consensus that ill effects from electromagnetic radiation have not been demonstrated, whereas the incidence of skin cancer is cause for serious concern. Related to this issue is that of immorality. Immorality play a role when technological solutions go against people’s ethical or moral principles.

Benefits can counterbalance the availability risks on the shortlist. Risky situations can be acceptable when the (perceived) benefits outweigh the (perceived) risks. For example, construction of a high broadcasting tower may meet with less opposition if it will be used for emergency communications instead of entertainment broadcasts.

Blame can sometime be apportioned to some actor (e.g. a telecoms operator), but natural risks cannot be blamed on anyone in particular. Risks without blame are often more acceptable than risk caused by some explicit actor.

Catastrophic potential makes risks less acceptable. The menace of wide-spread, large-scale destruction, regardless of likelihood, makes risks less tolerable. On the other hand, risks that have a small chronic effect over a period of time are often more easily accepted.

Children influence risk perception, sometimes in dramatic ways. People have strong feelings when children are affected by risks.

Familiarity with a risk may lead to complacency. The reverse is also true: novel risks may be less tolerable, simply because they are less well-known.

Fear is a general factor, related to catastrophic potential and familiarity. Strong feelings of fear decrease risk acceptance.

Institutional control can reassure people that risks are handled diligently. Sufficient trust in institutions is a prerequisite. When trust in institutions is low, risks will be perceived to be higher.

Media exposure can lead to increased perception of likelihood. Few people experience risks first-hand, and wide coverage by broadcasting or social networks can increase risk perception.

Mobilisation potential is relevant to decision makers. The ‘nimby’ phenomenon (“not in my backyard”) reflects mobilisation by nearby residents. Risks may provoke wide-spread and vocal opposition, making them less acceptable to decision makers.

Personal control refers to the amount of influence individuals can exert over the risky situation. For example, the risk of disturbances in communication are more acceptable when the user has the ability to control the device and participate in communication, instead of having only the passive ability to listen.

Violation of equity occurs when the benefits and the adverse effects are unevenly distributed. Opposition will be strong if the beneficiaries do not experience adverse effects at all.

Voluntariness is related to personal control. For example, people can choose whether or not to use a mobile phone, but construction of a mobile antenna mast in their neighbourhood is imposed upon them. Lack of voluntariness makes risks less acceptable.

6.3.3 Review the shortlist

The analysts must review each risk on the shortlist, to determine whether social risk factors may have a significant impact. This consists of the following steps:

1. Predict in what forms the risk factor would be expressed for various external stakeholders. For example, would it lead to a tarnished public image, reduced funding, or perhaps active opposition?
2. Assess the influence that this would have on the ability of decision makers to defend their choice of risk treatments. Can they easily deflect criticism, or will they be forced to select an alternative treatment?
3. Assess how the influence of the risk factor could be mitigated in advance, for example by informing stakeholders in advance, ask for their approval, or having them participate in a monitoring and oversight body.

If necessary, risk prioritisation should be adjusted and additional or different risk treatments should be recommended.

6.4 Prepare final report

The analysts have now collected all information for the final report. Not only can they present a prioritised shortlist of most serious risks with treatment recommendations, but they can also provide arguments for their proposals.

This final report must be reviewed by all analysts, and it must be approved by consensus before it is presented to the sponsor. The study is thereby concluded.

A suggested outline of the final report is shown below.

1. Executive summary to the final report.
2. About the crisis organisation (internal scope):
 - a. Position within wider system of crisis response.
 - b. Tasks.

- c. Responsibilities.
 - d. Telecom services used, together with their role and purpose during crisis response.
 - e. Main actors.
- 3. About the environment of the crisis organisation (external scope):
 - a. Disaster scenarios.
 - b. External parties with whom the main actors may communicate.
- 4. Roles and stakeholders
- 5. Telecom services
 - a. Diagram with explanation (once for each service)
 - b. Important risks (single failures and common cause failures)
- 6. Risk shortlist, with for each risk:
 - a. Description
 - b. Relevant social risk factors
 - c. Justification for risk priority, uncertainty, and limits to knowledge
 - d. Recommended risk treatment
- 7. Conclusions and recommended actions

Appendices:

- 8. Glossary
- 9. Reports of single failures
- 10. Reports of common cause failures

7 Executing the Raster method

Practical guidelines for execution of the Raster method.

7.1 Team composition

Three factors influence the choice and number of analysts.

1. To apply the Raster method to a crisis organisation, expertise from various fields of study is essential. Analysing threats to telecom service components requires in-depth knowledge of telecoms engineering, crisis management, political and legal issues, and the preferences of external stakeholders. No analyst can be expected to be expert in all these fields.
2. Raster requires analysts to make assessments about uncertain scenarios, often without access to all desired information. This inevitably means that assessments are partly subjective. By including several analysts from different backgrounds, the amount of subjectivity can be kept in check.
3. Several steps in the Raster method call for consensus. When the group becomes too large, reaching consensus will be time consuming.

These factors indicate that the group of analysts should not be too small, but also not too large. The group should include experts from different fields and backgrounds, and should not exceed 10 persons.

Some analysts may opt to not actively participate in the gathering of information and analysis of vulnerabilities. A core group of analysts will then perform most of the tasks. However, it is essential that all analysts participate in all reviews, and agree to the stage results and final report.

7.2 Managing work sessions

Before a Raster project can start, all analysts must be sufficiently familiar with the method. Each analyst should have received a copy of this manual well in advance. Unless all analysts are familiar with the method, an introductory session should be held in which someone who is well acquainted with the method shows its key activities using a small mock-example.

To speed up execution of the Raster method some activities can be performed in parallel. However, the more the analysts break into separate groups, the more they will need to coordinate the integration of their intermediate results later on. After all, the Raster method leads to a single final report, on which all analysts need to agree.

7.2.1 The recorder

During stages 2 and 3 one of the analysts should be appointed as recorder. The responsibility of the recorder is to record the diagrams and the assessments of vulnerabilities to components using the Raster tool. The recorder should use a computer connected to a projector, so that all analysts in the room can view a common, central display of the tool. The recorder may also act as a moderator during the assessment.

Because the recorder notes all assessments, he or she will be the best placed to detect inconsistencies in assessments. The recorder should take special care to notice inconsistent scores between components, and bring these up for discussion. For example, if some vulnerability is scored as Medium in one component but as Low in another, similar component, the group should discuss whether one of these scores may have to be adjusted.

In follow-up sessions the recorder may find it useful to distribute printouts from the Raster tool for reference (see section 13.3).

7.2.2 Stage 1 — Initiation and preparation

If some of the analysts are not yet familiar with the method, the group should not be divided. Otherwise, two groups could be formed:

1. one group to identify telecom services and actors, and
2. one group to describe disaster scenarios.

This division should only be made if the analysts expect that it will save a large amount of time. Because the Stage 1 results will be referenced throughout the study (in stages 2, 3, and 4) it is important that all analysts are intimately familiar with its contents, which may not be the case when the group is divided.

7.2.3 Stage 2 — Single failures analysis

For analysis of diagrams, the analysts may divide themselves into groups, each group analysing their own subset of telecom services. Before the group is split one or two services should first be analysed together, so that all analysts share a common approach and procedure. The groups must then remain in close contact, as components may be present in more than one telecom service and must be assessed in a consistent manner.

It may not be possible to complete the analysis in a single work session. After the initial diagram has been created and analysed, the analysts may decide to expand unknown links, or decide to do further investigations. It may thus require a number of work sessions to complete the single failures analysis.

7.2.4 Stage 3 — Common cause failures analysis

Common cause failures are assessed for the project as a whole, and not for individual telecom services. It is therefore not possible nor useful to divide the analysts into groups.

As with single failures analysis, multiple work sessions may be necessary to complete the analysis of common cause failures.

7.2.5 Stage 4 — Risk evaluation

All analysts must participate in creation of the longlist and shortlist. Most likely this can be completed in a single work session. Based on this shortlist social risk factors must be assessed. This assessment may require further investigation, and it may not be possible to complete this in a single work session. The analysts may decide to break into groups, each group analysing some risks from the shortlist. The first few risks should again be analysed together, before the group is split. As with all reviews, it is essential that the assessment of social risk factors be reviewed by the entire group of analysts.

The collation of material into a final report should be done by a small team of editors. Much of the Stage 1 report can be reused, and printouts from the Raster tool can be used for the appendices suggested in the template in section 6.4.

Quick reference

Frequency

<i>Class</i>	<i>Value</i>	<i>Symbol</i>
High	Once in 5 years. For 100 identical components, each month 1 or 2 will experience an incident.	H
Medium	Once in 50 years. For 100 identical components, each year 2 will experience an incident.	M
Low	Once in 500 years. For 100 identical components, one incident will occur every five years.	L
Extremely high	Routine event. Very often.	V
Extremely low	Very rare, but not physically impossible.	U
Ambiguous	Indicates lack of consensus between analysts.	A
Unknown	Indicates lack of knowledge or data.	X
Not yet analysed	Default. Indicates that no assessment has been done yet.	–

Impact

<i>Class</i>	<i>Value</i>	<i>Symbol</i>
High	Partial unavailability, if unrepairable. Total unavailability, if long-term.	H
Medium	Partial unavailability, if repairable (short-term or long-term). Total unavailability, if short-term.	M
Low	Noticeable degradation, repairable (short-term or long-term) or unrepairable.	L
Extremely high	Very long-term or unrepairable unavailability of the service.	V
Extremely low	Unnoticeable effects, or no actors affected.	U
Ambiguous	Indicates lack of consensus between analysts.	A
Unknown	Indicates lack of knowledge or data.	X
Not yet analysed	Default. Indicates that no assessment has been done yet.	–

Vulnerability levels

-	Not yet analysed, no assessment has been done yet (white)
A	Ambiguous, the assessors have conflicting opinions (purple)
V	Extremely (very) high, an extreme risk (bright red)
H	High (red)
M	Medium (yellow-orange)
L	Low (green-orange)
X	Unknown, because of lack of knowledge (sky blue).
U	Extremely (ultra, unlikely) low, the risk level is negligible or absent (bright green)

Overview of the Raster method

Stage 1 — Initiation and preparation	chapter 3
1. Identify telecom services	
2. Identify actors	
3. Describe disaster scenarios	
4. Create Stage 1 report	
5. Obtain approval from sponsor	
Stage 2 — Single failures analysis	chapter 4
1. Update the checklists of vulnerabilities	
2. Draw initial diagrams	
3. Analyse the frequency and impact of components	
4. Expand unknown links	
5. Review	
Stage 3 — Common cause failures analysis	chapter 5
1. Create clusters	
2. Analyse each cluster	
3. Expand unknown links	
4. Review	
Stage 4 — Risk evaluation	chapter 6
1. Determine longlist	
2. Reduce longlist to shortlist	
3. Draft treatment recommendations	
4. Assess social risk factors	
5. Prepare final report	

Suggested outline of the final report

1. Executive summary to the final report.
 2. About the crisis organisation (internal scope):
 - a) Position within wider system of crisis response.
 - b) Tasks.
 - c) Responsibilities.
 - d) Telecom services used, together with their role and purpose during crisis response.
 - e) Main actors.
 3. About the environment of the crisis organisation (external scope):
 - a) Disaster scenarios.
 - b) External parties with whom the main actors may communicate.
 4. Roles and stakeholders
 5. Telecom services
 - a) Diagram with explanation (once for each service)
 - b) Important risks (single failures and common cause failures)
 6. Risk shortlist, with for each risk:
 - a) Description
 - b) Relevant social risk factors
 - c) Justification for risk priority
 - d) Recommended risk treatment
 7. Conclusions and recommended actions
- Appendices:
8. Glossary
 9. Reports of single failures
 10. Reports of common cause failures

References

Author publications

Refereed conference publications

1. Vriezekolk, E., Wieringa, R., and Etalle, S. (2015) "Validating the Raster Risk Assessment Method in Practice". In: Proceedings of the 12th International Conference on Information Systems for Crisis Response and Management ISCRAM 2015.
2. Vriezekolk, E., Wieringa, R., and Etalle, S. (2015) "Experimental validation of a risk assessment method". In: Proceedings of the 21st Working Conference on Requirements Engineering REFSQ 2015, LNCS, vol 9013, Springer.
3. Vriezekolk, E., Wieringa, R., and Etalle, S. (2012) "Design and initial validation of the Raster method for telecom service availability risk assessment". In: Proceedings of the 9th International Conference on Information Systems for Crisis Response and Management ISCRAM 2012.
4. Vriezekolk, E., Wieringa, R., and Etalle, S. (2011) "How to assess telecom service availability risks for crisis organisations?". Advances in Safety, Reliability and Risk Management: ESREL 2011, 2653-2661, Taylor & Francis.
5. Vriezekolk, E., Wieringa, R., and Etalle, S. (2011) "A new method to assess telecom service availability risks". In: Proceedings of the 8th International Conference on Information Systems for Crisis Response and Management ISCRAM 2011.

As co-author

6. Zarghami, A., Vriezekolk, E., Zarifi Eslami, M., van Sinderen, M.J. and Wieringa, P.A. (2013) "Assumption-based Risk Identification Method (ARM) in Dynamic Service Provisioning". In: Proceedings of the 21st IEEE International Requirements Engineering Conference 2013.
7. Zarifi Eslami, M., Sapkota, B., Zarghami, A., Vriezekolk, E., van Sinderen, M.J. and Wieringa, P.A. (2012) "Risk identification of tailorable context-aware systems: a case study and lessons learned". In: Proceedings of the CAiSE'12 Forum at the 24th International Conference on Advanced Information Systems Engineering (CAiSE) 2012.

Workshop and discussion papers

8. Vriezekolk, E. (2016) "Technical Report on RASTER field test at a Water Board". Technical Report TR-CTIT-16-06, Centre for Telematics and Information Technology, University of Twente, Enschede.

9. Vriezolk, E. (2015) "Worse than cybercrime: why availability matters". Presentation at the National Cyber Security Centre ONE Conference, The Hague, 2015.
10. Vriezolk, E. (2014) "Testing reliability of RASTER – report of experiment with Kerckhoffs students". Technical Report TR-CTIT-14-05, Centre for Telematics and Information Technology, University of Twente, Enschede.
11. Vriezolk, E., Wieringa, R., and Etalle, S. (2013) "Can STPA be used to investigate telecom incidents?". Presented at the First STAMP Workshop Europe 2013.

Other publications

12. Vriezolk, E. (2015) "The Raster Method – application manual". English version 1.0, 2015. Available at <https://github.com/EelcoV/RasterTool/>.
13. Vriezolk, E. (2015) "De Raster Methode – gebruikershandleiding". Version 1.0, 2015. Available at <https://github.com/EelcoV/RasterTool/>.

References

- [1] A. Abdulkhaleq and S. Wagner. Xstamp: An extensible stamp platform as tool support for safety engineering. In *2015 STAMP Workshop, MIT, Boston, USA*. Universität Stuttgart, 2015. URL <http://elib.uni-stuttgart.de/opus/volltexte/2015/9987>.
- [2] A. Ahlbom, A. Green, L. Kheifets, D. Savitz, and A. Swerdlow. Epidemiology of health effects of radiofrequency exposure. *Environmental Health Perspectives*, 112(17):1741, 2004.
- [3] D. Alberts, J. Garstka, and F. Stein. *Network centric warfare: Developing and leveraging information superiority*. CCRP publication series. National Defense University Press, 2nd edition (revised) edition, 1999.
- [4] G. Apostolakis. How useful is quantitative risk assessment? *Risk Analysis*, 24(3): 515–520, 2004.
- [5] T. Aven. *Foundations of Risk Analysis*. John Wiley & Sons, 2003.
- [6] T. Aven and V. Kristensen. Perspectives on risk: review and discussion of the basis for establishing a unified and holistic approach. *Reliability Engineering and System Safety*, 90(1):1 – 14, 2005.
- [7] T. Aven and O. Renn. On risk defined as an event where the outcome is uncertain. *Journal of Risk Research*, 12(1):1–11, 2009.
- [8] A. Avižienis, J.-C. Laprie, B. Randell, and C. Landwehr. Basic concepts and taxonomy of dependable and secure computing. *Dependable and Secure Computing, IEEE Transactions on*, 1(1):11–33, 2004.
- [9] B. Barber and J. Davey. The use of the CCTA risk analysis and management methodology CRAMM. In *Proc. MEDINFO92, North Holland*, volume 1593, 1992.
- [10] J. Barlow. A declaration of the independence of cyberspace. *The Humanist*, 56(3): 18–19, May/June 1996.
- [11] P. Bartha. *By Parallel Reasoning*. Oxford University Press, 2010.
- [12] U. Beck and M. Ritter. *Risk society: towards a new modernity*. Sage, 1992.
- [13] Berenschot and I&O Research. Continuïteitsplan Veiligheidsregio bij uitval van elektriciteit en/of ICT (in Dutch). Model plan, Ministry of Safety and Justice, April 2013.
- [14] B. Boehm, P. Bose, E. Horowitz, and M.-J. Lee. Software requirements as negotiated win conditions. In *Requirements Engineering, 1994., Proceedings of the First International Conference on*, pages 74–83, Apr 1994.
- [15] B. W. Boehm. Software risk management: principles and practices. *IEEE Software*, 8(1):32–41, Jan 1991.
- [16] D. Borys, D. Else, and S. Legget. The fifth age of safety: the adaptive age. *Journal of Health & Safety research & Practice*, 1(1), 2009.
- [17] J. Bradbury. The policy implications of differing concepts of risk. *Science, Technology & Human Values*, 14(4):380, 1989.
- [18] D. Carpenter, C. Sage, et al. Bioinitiative Report. A Rationale for a Biologically-based Public Exposure Standard for Electromagnetic Fields (ELF and RF). Technical report, A collaboration of international scientists, 2007.

- [19] R. Carson. *Silent Spring*. Houghton Mifflin, 1962.
- [20] M. Cauvin, R. Salmon, and T. Verdel. Dealing with uncertainties in the context of post mining hazard evaluation. In *Symposium Post-Mining 2008*. ASGA. Vandoeuvres-Nancy, 2008.
- [21] F. Christensen, O. Andersen, N. Duijm, and P. Harremoës. Risk terminology – a platform for common understanding and better communication. *Journal of hazardous materials*, 103(3):181–203, 2003.
- [22] J. J. Cochran, L. A. Cox, P. Keskinocak, J. P. Kharoufeh, and J. C. Smith. *MACBETH (Measuring Attractiveness by a Categorical Based Evaluation Technique)*. Wiley, 2010.
- [23] M. Colyvan. Is probability the only coherent approach to uncertainty? *Risk Analysis*, 28(3):645–652, 2008.
- [24] R. M. Cooke and L. H. J. Goossens. Expert judgement elicitation for risk assessments of critical infrastructures. *Journal of Risk Research*, 7(6):643–656, 2004.
- [25] R. Cooke, L. Goossens, and B. Kraan. Probabilistic accident consequence uncertainty assessment procedures guide for structured expert judgement. Procedures Guide 18820, European commission, Directorate-General for Research, 2000.
- [26] S. Cooper, A. Ramey-Smith, J. Wreathall, and G. Parry. A technique for human error analysis (ATHEANA). Technical report, Nuclear Regulatory Commission, USA, 1996.
- [27] M. Cousin and M. Siegrist. Laypeople’s health concerns and health beliefs in regard to risk perception of mobile communication. *Human and Ecological Risk Assessment: An International Journal*, 14(6):1235–1249, 2008.
- [28] M. Cousin and M. Siegrist. Risk perception of mobile communication: a mental models approach. *Journal of Risk Research*, 13(5):599–620, 2010.
- [29] F. Cross. Facts and values in risk assessment. *Reliability Engineering and System Safety*, 59(1):27–40, 1998.
- [30] T. Das and B.-S. Teng. The risk-based view of trust: A conceptual framework. *Journal of Business and Psychology*, 19(1):85–116, 2004.
- [31] F. D. Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, pages 319–340, 1989.
- [32] B. De Finetti. *Theory of probability*. Wiley, 1974.
- [33] M. v. Dorssen, M. Holzmann, K. Franx, and L. Rens. Continuïteitsplannen ICT/elektriciteit: inventarisatie van continuïteitsplannen in geval van grootschalige uitval van ICT en/of elektriciteit bij gemeenten, provincies, veiligheidsregio’s, politieregio’s en waterschappen (in Dutch, with English summary). Technical report, Berenschot Groep/I&O Research, 2012. URL <http://wodc.nl/onderzoeksdatabase/stand-van-zaken-continuiteitsplannen-van-de-sectoren-ob-en-oov.aspx?cp=44&cs=6796>.
- [34] E. M. Dougherty. Human reliability analysis—where shouldst thou turn? *Reliability Engineering and System Safety*, 29(3):283 – 299, 1990.
- [35] J. Duffus, S. Brown, and N. Fernicola. Glossary for chemists of terms used in toxicology. *International Union of Pure and Applied Chemistry*, 65:2003–2122, 1993.
- [36] Dutch Scientific Council for Government Policy. *Onzekere veiligheid: verantwoordelijkheden rond fysieke veiligheid (in Dutch)*. Number 82 in WRR-rapportages. Amsterdam University Press, 2008. ISBN 978-90-5356-619-0.
- [37] European Commission. Guideline on the definition of a potential serious risk to

- public health in the context of Article 29(1) and (2) of Directive 2001/83/EC – March 2006. Communication C 133/5, European Union, 2006.
- [38] European Environment Agency. Environmental terminology and discovery service. Retrieved 2016-04-22. URL <http://glossary.eea.europa.eu/>.
 - [39] European Network and Information Security Agency ENISA. Risk management: Implementation principles and inventories for risk management/risk assessment methods and tools. Technical report, 2006.
 - [40] European Union. Treaty on European Union (The Maastricht Treaty). Office for Official Publications of the European Communities, 1992.
 - [41] M. S. Feather and S. L. Cornford. Quantitative risk-based requirements reasoning. *Requirements Engineering*, 8(4):248–265, 2003.
 - [42] S. Ferson and L. Ginzburg. Different methods are needed to propagate ignorance and variability. *Reliability Engineering and System Safety*, 54(2):133–144, 1996.
 - [43] B. Fischhoff, P. Slovic, S. Lichtenstein, S. Read, and B. Combs. How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits. *Policy Sciences*, 9(2):127–152, 1978.
 - [44] V. N. Franqueira, T. T. Tun, Y. Yu, R. Wieringa, and B. Nuseibeh. Risk and argument: a risk-based argumentation method for practical security. In *Requirements Engineering Conference*, pages 239–248. IEEE, 2011.
 - [45] R. Fredriksen, M. Kristiansen, B. Gran, K. Stølen, T. Opperud, and T. Dimitrakos. The CORAS framework for a model-based risk management process. In *Proceedings of the 21st International Conference on Computer Safety, Reliability and Security*, pages 39–53. Springer, 2002.
 - [46] R. Frigg and S. Hartmann. Models in science. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Fall 2012 edition, 2012.
 - [47] D. Gardner. *Risk: The Science and politics of fear*. Virgin Books, 2009.
 - [48] German Advisory Council on Global Change (WBGU). *World in Transition: Strategies for Managing Global Environmental Risks*. Springer, 2000.
 - [49] S. Ghaisas, P. Rose, M. Daneva, and K. Sikkell. Generalizing by similarity: Lessons learnt from industrial case studies. In *1st International Workshop on Conducting Empirical Studies in Industry (CESI)*, pages 37–42, 2013.
 - [50] G. Gigerenzer, Z. Swijtink, T. Porter, L. Daston, J. Beatty, and L. Kruger. *The empire of chance: How probability changed science and everyday life*. Cambridge University Press, 1989.
 - [51] A. I. Glendon, S. Clarke, and E. McKenna. *Human safety and risk management*. CRC Press, 2006.
 - [52] Global Conference on Cyberspace. The Hague Declaration on the Global Forum on Cyber Expertise. Retrieved 2016-04-22, 16 April 2015. URL <https://www.gccs2015.com/gfce>.
 - [53] J. Goldsmith and T. Wu. *Who Controls the Internet? Illusions of a Borderless World*. Oxford University Press, 2006.
 - [54] T. Gorschek, P. Garre, S. Larsson, and C. Wohlin. A model for technology transfer in practice. *IEEE Software*, 23(6):88–95, 2006.
 - [55] I. Hacking. *An introduction to probability and inductive logic*. Cambridge University Press, 2001.
 - [56] A. Hale and J. Hovden. *Occupational injury: Risk, prevention and intervention*, chapter

- Management and culture: the third age of safety. A review of approaches to organizational aspects of safety, health and environment, pages 129–165. Taylor & Francis, London, 1998.
- [57] K. A. Hallgren. Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in quantitative methods for psychology*, 8(1):23, 2012.
 - [58] S. Hansson. The epistemology of technological risk. *Techné: Research in Philosophy and Technology*, 9(2), 2005.
 - [59] S. Hansson. Risk: objective or subjective, facts or values. *Journal of Risk Research*, 13(2):231–238, 2010.
 - [60] Health Council of the Netherlands. Not all risks are equal. A commentary on ‘Premises for Risk Management’. Report 1995/06E, Health Council of the Netherlands, The Hague, 1995.
 - [61] Health Council of the Netherlands. Electromagnetic fields: Annual update 2008. Technical Report 2009/02, Health Council of the Netherlands, 2008.
 - [62] I. Helsloot. *Voorbij de symboliek. Over de noodzaak van een rationeel perspectief op fysiek veiligheidsbeleid (in Dutch)*. Boom, 2007.
 - [63] J. Helton and W. Oberkampf. Alternative representations of epistemic uncertainty. *Reliability Engineering and System Safety*, 85(1-3):1–10, 2004.
 - [64] A. Herrmann. Information need of IT risk estimation – qualitative results from experiments. In *Proceedings of the REFSQ 2011 RePriCo Workshop*, pages 72–84, 2011.
 - [65] A. Herrmann. REFSQ 2011 live experiment about risk-based requirements prioritization: The influence of wording and metrics. In *Proceedings of REFSQ 2011*, pages 176–194, 2011.
 - [66] A. Herrmann and B. Paech. Practical challenges of requirements prioritization based on risk estimation. *Empirical Software Engineering*, 14(6):644–684, 2009.
 - [67] F. O. Hoffman and J. S. Hammonds. Propagation of uncertainty in risk assessments: the need to distinguish between uncertainty due to lack of knowledge and uncertainty due to variability. *Risk analysis*, 14(5):707–712, 1994.
 - [68] A. Hollander and A. Hanemaaijer. Nuchter omgaan met risico’s (in Dutch)). RIVM rapport 251701047/2003, Milieu- en Natuurplanbureau, Rijksinstituut voor Volksgezondheid en Milieu, 2003.
 - [69] E. Hollnagel and O. Goteman. The functional resonance accident model. *Cognitive System Engineering in Process Plant 2004*, 2004.
 - [70] E. Hollnagel. *Cognitive reliability and error analysis method (CREAM)*. Elsevier, 1998.
 - [71] M. Höst, B. Regnell, and C. Wohlin. Using students as subjects – a comparative study of students and professionals in lead-time impact assessment. *Empirical Software Engineering*, 5(3):201–214, 2000.
 - [72] P. Hudson. Accident causation models, management and the law. *Journal of Risk Research*, 17(6):749–764, 2014.
 - [73] Ingenieurs/Adviesbureau SAVE & Adviesbureau Van Dijke. Leidraad maatramp. Ministry of Internal Affairs, version 1.3 2000.
 - [74] Inspectorate of Security and Justice. Storing telecommunicatienetwerk Waalhaven Rotterdam (in Dutch). Investigation Report J-14452, Ministry of Safety and Justice, Netherlands, 2012.
 - [75] Inspectorate of Security and Justice. 1-1-2 onder de loep (in Dutch). Investigation Report J-18240, Ministry of Safety and Justice, Netherlands, 2013.

- [76] International Organization for Standardization. Information technology – security techniques – guidelines for the management of IT security – part 1: Concepts and models for it security. International standard, ISO/IEC, 13335-1:2004.
- [77] International Organization for Standardization. Information technology – security techniques – evaluation criteria for IT security – part 1: Introduction and general model. International standard, ISO/IEC, 15408-1:2005.
- [78] International Organization for Standardization. Information technology – Open Distributed Processing – Unified Modeling Language (UML) Version 1.4.2. International standard, ISO, 19501:2005.
- [79] International Organization for Standardization. Societal security – guideline for incident preparedness and operational continuity management. International standard, ISO/PAS, 22399:2007.
- [80] International Organization for Standardization. Information technology – security techniques – information security management systems - overview and vocabulary. International standard, ISO/IEC, 27000:2014.
- [81] International Organization for Standardization. Information technology – security techniques – code of practice for information security controls. International standard, ISO/IEC, 27002:2013.
- [82] International Organization for Standardization. Information technology – security techniques – information security risk management. International standard, ISO/IEC, 27005:2011.
- [83] International Organization for Standardization. Information technology – security techniques – management guidelines for telecommunications organizations based on ISO/IEC 27002. International standard, ISO/IEC, 27011:2008.
- [84] International Organization for Standardization. Information technology – security techniques – information security incident management. International standard, ISO/IEC, 27035:2011.
- [85] International Organization for Standardization. Risk management – principles and guidelines. International standard, ISO, 31000:2009.
- [86] International Organization for Standardization. Risk management – risk assessment techniques. International standard, ISO/IEC, 31010:2009.
- [87] International Organization for Standardization. Analysis techniques for system reliability – Procedure for failure mode and effects analysis (FMEA). International standard, 60812:2006.
- [88] International Organization for Standardization. Fault tree analysis (FTA). International standard, 61025:2006.
- [89] International Organization for Standardization. Hazard and operability studies (HAZOP studies) – application guide. International standard, 61882:2001.
- [90] International Organization for Standardization. Risk management – vocabulary. Guide, ISO/IEC, 73:2009.
- [91] W. Johnson. Management oversight and risk tree – MORT. Technical report, Aerojet Nuclear Company, USA, 1973.
- [92] W. Jong and R. Johannink. *Loco voor de leeuwen – crisiservaringen van loco-burgemeesters (in Dutch)*. Nederlands Genootschap van Burgemeesters, 2009.
- [93] D. Kahneman. *Thinking, fast and slow*. Macmillan, 2011.
- [94] S. Kaplan and B. Garrick. On the quantitative definition of risk. *Risk Analysis*, 1(1): 11–27, 1981.

- [95] R. Kasperson, O. Renn, P. Slovic, H. Brown, J. Emel, R. Goble, J. Kasperson, and S. Ratick. The social amplification of risk: A conceptual framework. *Risk Analysis*, 8(2):177–187, 1988.
- [96] Kerckhoffs Institute. The Kerckhoffs masters programme. Retrieved 2016-04-22. URL <http://www.kerckhoffs-institute.org/>.
- [97] A. Klinke and O. Renn. A new approach to risk evaluation and management: Risk-based, precaution-based, and discourse-based strategies. *Risk Analysis*, 22(6): 1071–1094, 2002.
- [98] K. Krippendorff. Calculation of alpha over partitions. Private communication.
- [99] K. Krippendorff. *Content analysis: an introduction to its methodology*. Sage, 2nd edition, 2004.
- [100] V. Kristensen, T. Aven, and D. Ford. A new perspective on Renn and Klinke’s approach to risk evaluation and management. *Reliability Engineering and System Safety*, 91(4):421–432, 2006.
- [101] L1 Radio. A2-matrixborden van slag tijdens werkzaamheden (in Dutch). Retrieved 2016-04-22. URL <http://www.l1.nl/nieuws/a2-matrixborden-van-slag-tijdens-werkzaamheden>.
- [102] T. R. LaPorte and P. M. Consolini. Working in practice but not in theory: Theoretical challenges of “High-Reliability Organization”. *Journal of Public Administration Research and Theory*, 1(1):19–48, 1991.
- [103] P. Le Bot, E. Desmares, C. Bieder, J. Bonnet, and F. Cara. MERMOS. An EDF project to update the PHRA methodology. Technical report, Electricité de France, 1997.
- [104] W. Leiss. "Down and Dirty:" the use and abuse of public trust in risk communication. *Risk Analysis*, 15(6):685–692, 1995.
- [105] N. Leveson. A new accident model for engineering safer systems. *Safety Science*, 42(4):237–270, 2004.
- [106] N. Leveson. An STPA primer. Technical report, MIT Partnership for a Systems Approach to Safety (PSAS), <http://psas.scripts.mit.edu/home/>, 2015.
- [107] N. Leveson. *Safeware: system safety and computers*. Addison-Wesley, 1995.
- [108] N. Leveson. *Engineering a safer world: Systems thinking applied to safety*. MIT Press, 2011.
- [109] N. Leveson. STPA and CAST tutorial. In *First STAMP/STPA workshop at MIT*. <http://psas.scripts.mit.edu/home/1st-stampstpa-workshop-2012/>, 2012.
- [110] N. Leveson, N. Dulac, K. Marais, and J. Carroll. Moving beyond normal accidents and high reliability organizations: A systems approach to safety in complex systems. *Organization Studies*, 30(2–3):227–249, February/March 2009.
- [111] M. S. Lund, B. Solhaug, and K. Stølen. *Model-driven risk analysis: the CORAS approach*. Springer, 2010.
- [112] E. Muller, U. Rosenthal, I. Helsloot, and E. van Dijkman. *Crisis: studies over crisis en crisisbeheersing (in Dutch)*. Kluwer, 2009.
- [113] Netherlands, Ministry of the Interior and Kingdom Relations. Veiligheidsrapportage 1993 (in Dutch). Parliamentary document 23.096/2, Tweede Kamer der Staten-Generaal, 1993.
- [114] Netherlands, Parliament. New infrastructure mobile communication C2000 (in Dutch). House document 25124 nr 1, 1996.

- [115] Netherlands, Parliament. Besluit veiligheidsregio's (in Dutch). Ministerieel besluit, 2009.
- [116] Netherlands, Parliament. Wet veiligheidsregio's, Memorie van Toelichting (in Dutch). Technical report, 2009.
- [117] Netherlands, Parliament. New infrastructure mobile communication C2000 (in Dutch). House document 25124 nr 64, 2010.
- [118] NRC Handelsblad. Vodafone heeft storing niet onder controle (in Dutch). Retrieved 2016-04-22. URL <http://vorige.nrc.nl/economie/article2417570.ece>.
- [119] Ofcom. The Consumer Experience. Annex 6. Consumer decision making in the telecoms and TV broadcast markets. Research report 7, United Kingdom, Office of Communications, 2007.
- [120] Official Journal of the European Communities. On the control of major-accident hazards involving dangerous substances. EU Directive 82, European Union, 1996.
- [121] Official Journal of the European Communities. Establishing the European Network and Information Security Agency. EU Directive 460/2004, European Union, 2004.
- [122] Official Journal of the European Communities. On the European Environment Agency and the European Environment Information and Observation Network. EU Directive 401/2009, European Union, 2009.
- [123] B. Ojameruaye and R. Bahsoon. Systematic elaboration of compliance requirements using compliance debt and portfolio theory. In C. Salinesi and I. van de Weerd, editors, *Requirements Engineering: Foundation for Software Quality*, volume 8396 of *Lecture Notes in Computer Science*, pages 152–167. Springer, 2014.
- [124] T. Ord, R. Hillerbrand, and A. Sandberg. Probing the improbable: methodological challenges for risks with low probabilities and high stakes. *Journal of Risk Research*, 13(2):191–205, 2010.
- [125] P. K. Oriogun. A survey of Boehm's work on the spiral models and COCOMO II—towards software development process quality improvement. *Software Quality Journal*, 8(1):53–62, 1999.
- [126] H. Otway and K. Thomas. Reflections on risk perception and policy. *Risk Analysis*, 2(2):69–82, 1982.
- [127] M. E. Paté-Cornell. Uncertainties in risk analysis: Six levels of treatment. *Reliability Engineering and System Safety*, 54(2-3):95 – 111, 1996.
- [128] C. Perrow. *Normal accidents: Living with high-risk technologies*. Princeton University Press, 1984.
- [129] Public Order and Safety Inspectorate. Poldercrash 25 februari 2009 (in Dutch, with English summary). Report, Netherlands, Ministry of the Interior and Kingdom Relations, 2009.
- [130] Public Order and Safety Inspectorate, Expert group C2000. Final report of the expert group C2000 (in Dutch). Report, Netherlands, Ministry of the Interior and Kingdom Relations, 2010.
- [131] Radiocommunications Agency Netherlands. Staat van de Ether 2009 (in Dutch). Yearly report, Agentschap Telecom, URL <http://www.agentschaptelecom.nl/>, 2009.
- [132] Radiocommunications Agency Netherlands. Staat van de Ether 2010 (in Dutch). Yearly report, Agentschap Telecom, URL <http://www.agentschaptelecom.nl/>, 2010.

- [133] Radiocommunications Agency Netherlands. Staat van de Ether 2013 (in Dutch). Yearly report, Agentschap Telecom, URL <http://www.agentschaptelecom.nl/>, 2013.
- [134] B. Rasmussen and K. Borch. Risk and science: are we moving into the fourth age of risk concerns? *Risk, Decision and Policy*, 8(1):1–15, 2003.
- [135] N. Rasmussen et al. Reactor Safety Study, WASH-1400. *NUREG*, 75:014, 1975.
- [136] J. Reason. The contribution of latent human failures to the breakdown of complex systems. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 327(1241):pp. 475–484, 1990.
- [137] O. Renn. The role of risk perception for risk management. *Reliability Engineering and System Safety*, 59(1):49–62, 1998.
- [138] M. D. Rogers. Scientific and technological uncertainty, the precautionary principle, scenarios and risk management. *Journal of Risk Research*, 4(1):1–15, 2001.
- [139] P. Runeson. Using students as experiment subjects—an analysis on graduate and freshmen student data. In *Proceedings of the 7th International Conference on Empirical Assessment in Software Engineering*, pages 95–102, 2003.
- [140] T. Saaty. How to make a decision: the analytic hierarchy process. *European journal of operational research*, 48(1):9–26, 1990.
- [141] W. Shadish, T. Cook, and D. Campbell. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin, 2002.
- [142] L. Sjöberg. Factors in risk perception. *Risk analysis*, 20(1):1–12, 2000.
- [143] P. Slovic. Trust, emotion, sex, politics, and science: Surveying the risk-assessment battlefield. *Risk Analysis*, 19(4):689–701, 1999.
- [144] P. Slovic, J. H. Flynn, and M. Layman. Perceived risk, trust, and the politics of nuclear waste. *Science*, 254(5038):1603–1607, 1991.
- [145] Society for Risk Analysis SRA. SRA Glossary. Retrieved 2016-04-22. URL <http://www.sra.org/resources>.
- [146] C. Som, L. Hilty, and T. Ruddy. The precautionary principle in the information society. *Human and Ecological Risk Assessment*, 10(5):787–799, 2004.
- [147] Standards Australia/Standards New Zealand. Risk management. Australian/New Zealand Standard 4360, 2004.
- [148] C. Sunstein. On analogical reasoning. *Harvard Law Review*, 106:741–790, 1993.
- [149] M. Svahnberg, A. Aurum, and C. Wohlin. Using students as subjects—an empirical evaluation. In *Proceedings of the Second ACM-IEEE international symposium on Empirical software engineering and measurement*, pages 288–290. ACM, 2008.
- [150] A. Swain and H. Guttman. Handbook of human-reliability analysis with emphasis on nuclear power plant applications. Final report. Technical report, Sandia National Labs, USA, 1983.
- [151] N. N. Taleb. *The black swan: The impact of the highly improbable fragility*. Penguin, 2008.
- [152] W. Ten Brinke, G. Saeijs, I. Helsloot, and J. Van Alphen. Safety chain approach in flood risk management. *Municipal Engineer*, 161(2):93–102, 2008.
- [153] TETRA and Critical Communications Association (TCCA). Terrestrial Trunked Radio. Retrieved 2016-04-22. URL <http://www.tandcca.com/>.
- [154] The Committee on Man and Radiation (COMAR). COMAR technical information statement: Expert reviews on potential health effects of radiofrequency electro-

- magnetic fields and comments on the BioInitiative report. *Health Physics*, 97(4), 2009.
- [155] W. Thomas and D. Thomas. *The child in America*. Knopf, 1928.
 - [156] K. Thywissen. Components of risk: a comparative glossary. SOURCE 2, UNU Institute for Environment and Human Security UNU-EHS, Bonn, Germany, 2006.
 - [157] A. Tversky and D. Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):pp. 1124–1131, 1974.
 - [158] United Nations. Rio Declaration on Environment and Development. Conference on Environment and Development (UNCED), 1992.
 - [159] United Nations International Strategy for Disaster Reduction UN-ISDR. Terminology on disaster risk reduction. Geneva, May 2009.
 - [160] M. van Asselt, W. Passchier, and M. Krayen von Kraus. Uncertainty assessment: An analysis of regulatory science on wireless communication technology, RF EMF and cancer risks. Report for the IMBA project, Maastricht University, 2009.
 - [161] J. van de Ven, R. van Rijk, P. Essens, and E. Frinking. Network centric operations in crisis management. In F. Fiedrich and B. V. de Walle, editors, *Proceedings of the 5th International ISCRAM Conference*, pages 764–773, Washington, DC, USA., 2008.
 - [162] C. Vlek. Judicious management of uncertain risks: I. Developments and criticisms of risk analysis and precautionary reasoning. *Journal of Risk Research*, (1):1–27, 2010.
 - [163] C. Vlek and P.-J. Stallen. Rational and personal aspects of risk. *Acta Psychologica*, 45 (1-3):273 – 300, 1980.
 - [164] R. J. Wieringa. Design science as nested problem solving. In *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology*, Philadelphia, pages 1–12, New York, 2009. ACM.
 - [165] R. Wieringa. *Design Science Methodology for Information Systems and Software Engineering*. Springer, 2014.
 - [166] J. Williams. A data-based method for assessing and reducing human error to improve operational performance. In *Human Factors and Power Plants, 1988., Conference Record for 1988 IEEE Fourth Conference on*, pages 436–450, 1988.
 - [167] P. Willner. Methods for assessing the validity of animal models of human psychopathology. In A. Boulton, G. Baker, and M. Martin-Iverson, editors, *Animal Models in Psychiatry, I*, volume 18 of *Neuromethods*, pages 1–23. Humana Press, 1991.
 - [168] World Health Organization. IPCS risk assessment terminology, 2004.
 - [169] World Health Organization. Electromagnetic fields and public health. Base stations and wireless technologies. Factsheet 304, World Health Organization, 2006.
 - [170] R. K. Yin. *Case study research: design and methods – fourth edition*. Sage, 2009.

Index

Note: the main definition of a term is indicated like this*. References to the RASTER manual in Appendix D are indicated by an *italic* page number.

- accident model, 32
 - domino, 32
 - swiss cheese, 32
 - systems theory, 33
- actor, 67, 77, 79, 217, 222
 - main, 222
 - secondary, 222
 - vulnerabilities, 225
- AHP, *see* Analytic Hierarchy Process
- aleatory uncertainty, 21*, 27
- ambiguity, 21
- ambiguous, 227, 229, 232, 234, 237
- ambiguous risks, 42
- analyst, 50, 112, 243
 - editor, 245
 - recorder, 244
- Analytic Hierarchy Process, 56
- artifact, 8
- artificiality, immorality, 54, 239
- assessor, 50
- asset, 16
- ATHEANA, 33
- backup, 43, 221, 238
- backward-logic, 36
- benefits, 54, 239
- black swan, 96
- blame, 54, 239
- C2000, 59
- catastrophe, 3
- catastrophic potential, 55, 239
- CCF, *see* common cause failure
- checklist, 26, 76, 225
- children, 55, 239
- class
 - ambiguous, 229, 232, 234, 237
 - extremely high, 229*, 231*, 237
 - extremely low, 229*, 231*
 - high, 229, 231, 237
 - low, 229, 231
 - medium, 229, 231
 - not yet analysed, 229
 - unknown, 229, 232, 234, 237
- classical risk analysis, 32, 42, 46, 53
- cluster, 86, 233
- cognitive bias, 27, 105
- common cause f. analysis stage, 78, 216, 233*, 245
- common cause failure, 20, 37, 78, 85, 212, 233
- complex risks, 42
- component, 70
- consistency, 101
- constructivism, 23, 56
- continuity plan framework, 43, 123
- CORAS, 36, 44*
- countermeasure, 228
- CRAMM, 36
- CREAM, 33
- crime, 222, 225
- crisis, 3
- crisis communication, 24
- crisis management, 4
- crisis organisation, 2, 212
- critical property, 86, 233
- decision maker, 50, 52, 53, 56, 211–213
- definite class, 72
- delayed effects, 54
- design science, 8
- diagram, 216, 226
- disaster, 3
- disaster scenario, 76, 222, 229
- editor, 245
- electromagnetic fields, 57
- emergency, 3
- EMF, *see* electromagnetic fields
- engineering cycle, 9
- environment, 16
- epistemic uncertainty, 21*, 27

- equipment, 68, 218, 226
- Event tree analysis, 32
- expansion, 70, 232, 234
- experts, 27
 - fallibility, 53
- external stakeholder, 213, 240
- extremely high, 229*, 231*, 237
- extremely low, 229*, 231*

- failure, 19
- Failure mode and effect analysis, 32, 37, 65
- fallback, 221, 238
- familiarity, 55, 239
- Fault tree analysis, 32, 37, 65
- fear, 55, 239
- FMEA, *see* Failure mode and effect analysis
- FMECA, *see* Failure mode and effect analysis
- formal risk, 50
- forward-logic, 36
- FRAM, *see* Functional resonance accident model
- frequency, 71*, 72, 96, 97, 227, 228*, 234
- frequentist probability, 27
- FTA, *see* Fault tree analysis
- Functional resonance accident model, 33, 65
- fuzzy logic, 27

- glossary, 222

- hazard, 17
- Hazard and operability studies, 32, 38
- HAZOP, *see* Hazard and operability studies
- HEART, 32
- high, 229, 231, 237
- high reliability organisations, 33

- impact, 71*, 73, 96, 97, 116, 227, 229*
- incident, 3
- indefinite class, 72
- informal risk, 50
- initiation and preparation stage, 76, 216, 221*, 244
- institutional control, 54, 239
- inter-rater reliability, 105
- ISO 27000 series, 36, 45

- Krippendorff's alpha, 106, 110, 165

- life time, 228
- longlist, 87, 237
- low, 229, 231

- Macbeth, 57
- Management oversight and risk tree, 32
- media, 55
- media exposure, 239
- medium, 229, 231
- MERMOS, 33
- mobilisation, 54, 239
- model, 16, 67
- monitoring, 228, 230
- MORT, *see* Management oversight and risk tree

- network centric operations, 6, 61
- nimby, 240
- node, 70
- normal accidents, 32

- objectivism, 23, 56
- opportunity, 19, 29
- overall vulnerability level, 75, 98

- persistency, 54, 116
- personal control, 54, 239
- precautionary principle, 35, 42
- preparation (risk treatment), 238
- preparation stage, *see* initiation and preparation stage
- prevention, 5, 238
- prioritise, 237
- pro-action, 5, 238
- probability, 27
 - frequentist, 27
 - subjective, 28
- project
 - create, 225

- quick win, 237

- RASTER tool, 85, 95, 207
- recorder, 244
- recovery, 6
- redundancy, 87, 230
- refinement, *see* expansion
- refinement stage, 77, 99, *see also* single failures analysis stage

- reliability, 18, 101
 - inter-rater, 105
- repeatability, 101
- reproducibility, 101
- requirements, 63
- response, 6
- reversibility, 54, 116
- review, 223, 232, 235, 241
- risk, 22–29
 - ambiguous, 42
 - complex, 42
 - formal, 50
 - informal, 50
 - positive effect, 19
 - simple, 42
 - uncertain, 42
- risk analysis, 26
 - classical, 53
- risk communication, 24
- risk evaluation, 28
- risk evaluation stage, 78, 216, 237*, 245
- risk factor, 51*, 53–56
 - social, 52
 - strong, 51
- risk governance, 34
- risk identification, 26
- risk management, 24
- risk mitigation, 29
- risk perception, 24
- risk prioritisation, 87
- risk target, 16
- risk treatment, 29, 238
- safety, 17, 18
 - emergent, 33
 - industrial, 31
 - internal, external, 18
- safety chain, 5
- safety culture, 33, 79
- safety region, 4, 43, 121
- security, 18
- shortlist, 87, 237
- simple risks, 42
- single failures analysis stage, 99, 216, 225, 244
 - common cause failures analysis, 216, 233*, 245
 - initiation and preparation, 216, 221*, 244
 - risk evaluation, 216, 237*, 245
 - single failures analysis, 216, 225, 244
- stage 1 report, 223
- stakeholder, 17, 50, 56
- STAMP, *see* Systems-Theoretic Accident Model and Processes
- STPA, *see* Systems-Theoretic Process Analysis
- subjective probability, 28
- Systems-Theoretic Accident Model and Process, 33, 38, 65
- Systems-Theoretic Process Analysis, 38, 100
- TAM, *see* Technology acceptance model
- target, 16
- technical action research, 122
- Technology acceptance model, 123, 129, 132, 141, 148, 200
- technology assessment, 34
- telecom service, 221
- telecom service models, 67
 - expansion, 70
 - limitations, 78
 - well-formed, 69, 95
- Terrestrial Trunked Radio, 59
- TETRA, *see* Terrestrial Trunked Radio
- THERP, 32
- threat, 17
 - positive effects, 19, 54
- trust, 55
- ubiquity, 54, 116
- uncertain risks, 42
- uncertainty, 20*, 27, 54
 - aleatory, 21*, 27
 - epistemic, 21*, 27
- unknown, 227, 229, 232, 234, 237
- unknown link, 218
 - expansion, 232, 234
- unknown links, 69
- validation
 - design, 80
- value (of assets), 17

violation of equity, 55, 239
voluntariness, 54, 239
vulnerability, 20
vulnerability level adjustment, 75, 86
vulnerability score, 73, 96

W_{BGU}, 35
W_{BGU} method, 40
wired link, 217, 226
wired links, 68
wireless link, 68, 218, 226

SIKS Dissertatiereeks

2009

- 2009-01 Rasa Jurgelenaite (RUN)
Symmetric Causal Independence Models
- 2009-02 Willem Robert van Hage (VU)
Evaluating Ontology-Alignment Techniques
- 2009-03 Hans Stol (UvT)
A Framework for Evidence-based Policy Making Using IT
- 2009-04 Josephine Nabukenya (RUN)
Improving the Quality of Organisational Policy Making using Collaboration Engineering
- 2009-05 Sietse Overbeek (RUN)
Bridging Supply and Demand for Knowledge Intensive Tasks - Based on Knowledge, Cognition, and Quality
- 2009-06 Muhammad Subianto (UU)
Understanding Classification
- 2009-07 Ronald Poppe (UT)
Discriminative Vision-Based Recovery and Recognition of Human Motion
- 2009-08 Volker Nannen (VU)
Evolutionary Agent-Based Policy Analysis in Dynamic Environments
- 2009-09 Benjamin Kanagwa (RUN)
Design, Discovery and Construction of Service-oriented Systems
- 2009-10 Jan Wielemaker (UVA)
Logic programming for knowledge-intensive interactive applications
- 2009-11 Alexander Boer (UVA)
Legal Theory, Sources of Law & the Semantic Web
- 2009-12 Peter Massuthé (TUE, Humboldt-Universitaet zu Berlin)
Operating Guidelines for Services
- 2009-13 Steven de Jong (UM)
Fairness in Multi-Agent Systems
- 2009-14 Maksym Korotkiy (VU)
From ontology-enabled services to service-enabled ontologies (making ontologies work in e-science with ONTO-SOA)
- 2009-15 Rinke Hoekstra (UVA)
Ontology Representation - Design Patterns and Ontologies that Make Sense
- 2009-16 Fritz Reul (UvT)
New Architectures in Computer Chess
- 2009-17 Laurens van der Maaten (UvT)
Feature Extraction from Visual Data
- 2009-18 Fabian Groffen (CWI)
Armada, An Evolving Database System
- 2009-19 Valentin Robu (CWI)
Modeling Preferences, Strategic Reasoning and Collaboration in Agent-Mediated Electronic Markets
- 2009-20 Bob van der Vecht (UU)
Adjustable Autonomy: Controlling Influences on Decision Making
- 2009-21 Stijn Vanderlooy (UM)
Ranking and Reliable Classification
- 2009-22 Pavel Serdyukov (UT)
Search For Expertise: Going beyond direct evidence
- 2009-23 Peter Hofgesang (VU)
Modelling Web Usage in a Changing Environment
- 2009-24 Annerieke Heuvelink (VUA)
Cognitive Models for Training Simulations
- 2009-25 Alex van Ballegooij (CWI)
RAM: Array Database Management through Relational Mapping
- 2009-26 Fernando Koch (UU)
An Agent-Based Model for the Development of Intelligent Mobile Services
- 2009-27 Christian Glahn (OU)
Contextual Support of social Engagement and Reflection on the Web
- 2009-28 Sander Evers (UT)
Sensor Data Management with Probabilistic Models
- 2009-29 Stanislav Pokraev (UT)
Model-Driven Semantic Integration of Service-Oriented Applications
- 2009-30 Marcin Zukowski (CWI)
Balancing vectorized query execution with bandwidth-optimized storage
- 2009-31 Sofiya Katrenko (UVA)
A Closer Look at Learning Relations from Text
- 2009-32 Rik Farenhorst (VU) and Remco de Boer (VU)
Architectural Knowledge Management: Supporting Architects and Auditors
- 2009-33 Khiat Truong (UT)
How Does Real Affect Affect Recognition In Speech?
- 2009-34 Inge van de Weerd (UU)
Advancing in Software Product Management: An Incremental Method Engineering Approach
- 2009-35 Wouter Koelewijn (UL)
Privacy en Politiegegevens; Over geautomatiseerde normatieve informatie-uitwisseling
- 2009-36 Marco Kalz (OUN)
Placement Support for Learners in Learning Networks

- 2009-37 Hendrik Drachsler (OUN)
Navigation Support for Learners in Informal Learning Networks
- 2009-38 Riina Vuorikari (OU)
Tags and self-organisation: a metadata ecology for learning resources in a multilingual context
- 2009-39 Christian Stahl (TUE, Humboldt-Universitaet zu Berlin)
Service Substitution – A Behavioral Approach Based on Petri Nets
- 2009-40 Stephan Raaijmakers (UvT)
Multinomial Language Learning: Investigations into the Geometry of Language
- 2009-41 Igor Berezhnnyy (UvT)
Digital Analysis of Paintings
- 2009-42 Toine Bogers (UvT)
Recommender Systems for Social Bookmarking
- 2009-43 Virginia Nunes Leal Franqueira (UT)
Finding Multi-step Attacks in Computer Networks using Heuristic Search and Mobile Ambients
- 2009-44 Roberto Santana Tapia (UT)
Assessing Business-IT Alignment in Networked Organizations
- 2009-45 Jilles Vreeken (UU)
Making Pattern Mining Useful
- 2009-46 Loredana Afanasiev (UvA)
Querying XML: Benchmarks and Recursion
- ## 2010
- 2010-01 Matthijs van Leeuwen (UU)
Patterns that Matter
- 2010-02 Ingo Wassink (UT)
Work flows in Life Science
- 2010-03 Joost Geurts (CWI)
A Document Engineering Model and Processing Framework for Multimedia documents
- 2010-04 Olga Kulyk (UT)
Do You Know What I Know? Situational Awareness of Co-located Teams in Multidisplay Environments
- 2010-05 Claudia Hauff (UT)
Predicting the Effectiveness of Queries and Retrieval Systems
- 2010-06 Sander Bakkes (UvT)
Rapid Adaptation of Video Game AI
- 2010-07 Wim Fikkert (UT)
Gesture interaction at a Distance
- 2010-08 Krzysztof Siewicz (UL)
Towards an Improved Regulatory Framework of Free Software. Protecting user freedoms in a world of software communities and eGovernments
- 2010-09 Hugo Kielman (UL)
A Politiele gegevensverwerking en Privacy, Naar een effectieve waarborging
- 2010-10 Rebecca Ong (UL)
Mobile Communication and Protection of Children
- 2010-11 Adriaan Ter Mors (TUD)
The world according to MARP: Multi-Agent Route Planning
- 2010-12 Susan van den Braak (UU)
Sensemaking software for crime analysis
- 2010-13 Gianluigi Folino (RUN)
High Performance Data Mining using Bio-inspired techniques
- 2010-14 Sander van Splunter (VU)
Automated Web Service Reconfiguration
- 2010-15 Lianne Bodestaff (UT)
Managing Dependency Relations in Inter-Organizational Models
- 2010-16 Sicco Verwer (TUD)
Efficient Identification of Timed Automata, theory and practice
- 2010-17 Spyros Kotoulas (VU)
Scalable Discovery of Networked Resources: Algorithms, Infrastructure, Applications
- 2010-18 Charlotte Gerritsen (VU)
Caught in the Act: Investigating Crime by Agent-Based Simulation
- 2010-19 Henriette Cramer (UvA)
People's Responses to Autonomous and Adaptive Systems
- 2010-20 Ivo Swartjes (UT)
Whose Story Is It Anyway? How Improv Informs Agency and Authorship of Emergent Narrative
- 2010-21 Harold van Heerde (UT)
Privacy-aware data management by means of data degradation
- 2010-22 Michiel Hildebrand (CWI)
End-user Support for Access to Heterogeneous Linked Data
- 2010-23 Bas Steunebrink (UU)
The Logical Structure of Emotions
- 2010-24 Dmytro Tykhonov
Designing Generic and Efficient Negotiation Strategies
- 2010-25 Zulfiqar Ali Memon (VU)
Modelling Human-Awareness for Ambient Agents: A Human Mindreading Perspective
- 2010-26 Ying Zhang (CWI)
XRPC: Efficient Distributed Query Processing on Heterogeneous XQuery Engines
- 2010-27 Marten Voulon (UL)
Automatisch contracteren
- 2010-28 Arne Koopman (UU)
Characteristic Relational Patterns
- 2010-29 Stratos Idreos (CWI)
Database Cracking: Towards Auto-tuning Database Kernels
- 2010-30 Marieke van Erp (UvT)
Accessing Natural History - Discoveries in data cleaning, structuring, and retrieval
- 2010-31 Victor de Boer (UVA)
Ontology Enrichment from Heterogeneous Sources on the Web

- 2010-32 Marcel Hiel (UvT)
An Adaptive Service Oriented Architecture: Automatically solving Interoperability Problems
- 2010-33 Robin Aly (UT)
Modeling Representation Uncertainty in Concept-Based Multimedia Retrieval
- 2010-34 Teduh Dirgahayu (UT)
Interaction Design in Service Compositions
- 2010-35 Dolf Trieschnigg (UT)
Proof of Concept: Concept-based Biomedical Information Retrieval
- 2010-36 Jose Janssen (OU)
Paving the Way for Lifelong Learning; Facilitating competence development through a learning path specification
- 2010-37 Niels Lohmann (TUE)
Correctness of services and their composition
- 2010-38 Dirk Fahland (TUE)
From Scenarios to components
- 2010-39 Ghazanfar Farooq Siddiqui (VU)
Integrative modeling of emotions in virtual agents
- 2010-40 Mark van Assem (VU)
Converting and Integrating Vocabularies for the Semantic Web
- 2010-41 Guillaume Chaslot (UM)
Monte-Carlo Tree Search
- 2010-42 Sybren de Kinderen (VU)
Needs-driven service bundling in a multi-supplier setting - the computational e3-service approach
- 2010-43 Peter van Kranenburg (UU)
A Computational Approach to Content-Based Retrieval of Folk Song Melodies
- 2010-44 Pieter Bellekens (TUE)
An Approach towards Context-sensitive and User-adapted Access to Heterogeneous Data Sources, Illustrated in the Television Domain
- 2010-45 Vasilios Andrikopoulos (UvT)
A theory and model for the evolution of software services
- 2010-46 Vincent Pijpers (VU)
e3alignment: Exploring Inter-Organizational Business-ICT Alignment
- 2010-47 Chen Li (UT)
Mining Process Model Variants: Challenges, Techniques, Examples
- 2010-48 Withdrawn
- 2010-49 Jahn-Takeshi Saito (UM)
Solving difficult game positions
- 2010-50 Bouke Huurnink (UVA)
Search in Audiovisual Broadcast Archives
- 2010-51 Alia Khairia Amin (CWI)
Understanding and supporting information seeking tasks in multiple sources
- 2010-52 Peter-Paul van Maanen (VU)
Adaptive Support for Human-Computer Teams: Exploring the Use of Cognitive Models of Trust and Attention
- 2010-53 Edgar Meij (UVA)
Combining Concepts and Language Models for Information Access
- ## 2011
- 2011-01 Botond Cseke (RUN)
Variational Algorithms for Bayesian Inference in Latent Gaussian Models
- 2011-02 Nick Tinnemeier (UU)
Organizing Agent Organizations. Syntax and Operational Semantics of an Organization-Oriented Programming Language
- 2011-03 Jan Martijn van der Werf (TUE)
Compositional Design and Verification of Component-Based Information Systems
- 2011-04 Hado van Hasselt (UU)
Insights in Reinforcement Learning; Formal analysis and empirical evaluation of temporal-difference learning algorithms
- 2011-05 Base van der Raadt (VU)
Enterprise Architecture Coming of Age - Increasing the Performance of an Emerging Discipline.
- 2011-06 Yiwen Wang (TUE)
Semantically-Enhanced Recommendations in Cultural Heritage
- 2011-07 Yujia Cao (UT)
Multimodal Information Presentation for High Load Human Computer Interaction
- 2011-08 Nieske Vergunst (UU)
BDI-based Generation of Robust Task-Oriented Dialogues
- 2011-09 Tim de Jong (OU)
Contextualised Mobile Media for Learning
- 2011-10 Bart Bogaert (UvT)
Cloud Content Contention
- 2011-11 Dhaval Vyas (UT)
Designing for Awareness: An Experience-focused HCI Perspective
- 2011-12 Carmen Bratosin (TUE)
Grid Architecture for Distributed Process Mining
- 2011-13 Xiaoyu Mao (UvT)
Airport under Control. Multiagent Scheduling for Airport Ground Handling
- 2011-14 Milan Lovric (EUR)
Behavioral Finance and Agent-Based Artificial Markets
- 2011-15 Marijn Koolen (UvA)
The Meaning of Structure: the Value of Link Evidence for Information Retrieval
- 2011-16 Maarten Schadd (UM)
Selective Search in Games of Different Complexity
- 2011-17 Jiyin He (UVA)
Exploring Topic Structure: Coherence, Diversity and Relatedness
- 2011-18 Mark Ponsen (UM)
Strategic Decision-Making in complex games
- 2011-19 Ellen Rusman (OU)
The Mind's Eye on Personal Profiles

- 2011-20 Qing Gu (VU)
Guiding service-oriented software engineering - A view-based approach
- 2011-21 Linda Terlouw (TUD)
Modularization and Specification of Service-Oriented Systems
- 2011-22 Junte Zhang (UVA)
System Evaluation of Archival Description and Access
- 2011-23 Wouter Weerkamp (UVA)
Finding People and their Utterances in Social Media
- 2011-24 Herwin van Welbergen (UT)
Behavior Generation for Interpersonal Coordination with Virtual Humans On Specifying, Scheduling and Realizing Multimodal Virtual Human Behavior
- 2011-25 Syed Waqar ul Qounain Jaffry (VU)
Analysis and Validation of Models for Trust Dynamics
- 2011-26 Matthijs Aart Pontier (VU)
Virtual Agents for Human Communication - Emotion Regulation and Involvement-Distance Trade-Offs in Embodied Conversational Agents and Robots
- 2011-27 Aniel Bhulai (VU)
Dynamic website optimization through autonomous management of design patterns
- 2011-28 Rianne Kaptein (UVA)
Effective Focused Retrieval by Exploiting Query Context and Document Structure
- 2011-29 Faisal Kamiran (TUE)
Discrimination-aware Classification
- 2011-30 Egon van den Broek (UT)
Affective Signal Processing (ASP): Unraveling the mystery of emotions
- 2011-31 Ludo Waltman (EUR)
Computational and Game-Theoretic Approaches for Modeling Bounded Rationality
- 2011-32 Nees-Jan van Eck (EUR)
Methodological Advances in Bibliometric Mapping of Science
- 2011-33 Tom van der Weide (UU)
Arguing to Motivate Decisions
- 2011-34 Paolo Turrini (UU)
Strategic Reasoning in Interdependence: Logical and Game-theoretical Investigations
- 2011-35 Maaïke Harbers (UU)
Explaining Agent Behavior in Virtual Training
- 2011-36 Erik van der Spek (UU)
Experiments in serious game design: a cognitive approach
- 2011-37 Adriana Burlutiu (RUN)
Machine Learning for Pairwise Data, Applications for Preference Learning and Supervised Network Inference
- 2011-38 Nyree Lemmens (UM)
Bee-inspired Distributed Optimization
- 2011-39 Joost Westra (UU)
Organizing Adaptation using Agents in Serious Games
- 2011-40 Viktor Clerc (VU)
Architectural Knowledge Management in Global Software Development
- 2011-41 Luan Ibraimi (UT)
Cryptographically Enforced Distributed Data Access Control
- 2011-42 Michal Sindlar (UU)
Explaining Behavior through Mental State Attribution
- 2011-43 Henk van der Schuur (UU)
Process Improvement through Software Operation Knowledge
- 2011-44 Boris Reuderink (UT)
Robust Brain-Computer Interfaces
- 2011-45 Herman Stehouwer (UvT)
Statistical Language Models for Alternative Sequence Selection
- 2011-46 Beibei Hu (TUD)
Towards Contextualized Information Delivery: A Rule-based Architecture for the Domain of Mobile Police Work
- 2011-47 Azizi Bin Ab Aziz (VU)
Exploring Computational Models for Intelligent Support of Persons with Depression
- 2011-48 Mark Ter Maat (UT)
Response Selection and Turn-taking for a Sensitive Artificial Listening Agent
- 2011-49 Andreea Niculescu (UT)
Conversational interfaces for task-oriented spoken dialogues: design aspects influencing interaction quality

2012

- 2012-01 Terry Kakeeto (UvT)
Relationship Marketing for SMEs in Uganda
- 2012-02 Muhammad Umair (VU)
Adaptivity, emotion, and Rationality in Human and Ambient Agent Models
- 2012-03 Adam Vanya (VU)
Supporting Architecture Evolution by Mining Software Repositories
- 2012-04 Jurriaan Souer (UU)
Development of Content Management System-based Web Applications
- 2012-05 Marijn Plomp (UU)
Maturing Interorganisational Information Systems
- 2012-06 Wolfgang Reinhardt (OU)
Awareness Support for Knowledge Workers in Research Networks
- 2012-07 Rianne van Lambalgen (VU)
When the Going Gets Tough: Exploring Agent-based Models of Human Performance under Demanding Conditions
- 2012-08 Gerben de Vries (UVA)
Kernel Methods for Vessel Trajectories

- 2012-09 Ricardo Neisse (UT)
Trust and Privacy Management Support for Context-Aware Service Platforms
- 2012-10 David Smits (TUE)
Towards a Generic Distributed Adaptive Hyper-media Environment
- 2012-11 J.C.B. Rantham Prabhakara (TUE)
Process Mining in the Large: Preprocessing, Discovery, and Diagnostics
- 2012-12 Kees van der Sluijs (TUE)
Model Driven Design and Data Integration in Semantic Web Information Systems
- 2012-13 Suleman Shahid (UvT)
Fun and Face: Exploring non-verbal expressions of emotion during playful interactions
- 2012-14 Evgeny Knutov (TUE)
Generic Adaptation Framework for Unifying Adaptive Web-based Systems
- 2012-15 Natalie van der Wal (VU)
Social Agents. Agent-Based Modelling of Integrated Internal and Social Dynamics of Cognitive and Affective Processes.
- 2012-16 Fiemke Both (VU)
Helping people by understanding them - Ambient Agents supporting task execution and depression treatment
- 2012-17 Amal Elgammal (UvT)
Towards a Comprehensive Framework for Business Process Compliance
- 2012-18 Eltjo Poort (VU)
Improving Solution Architecting Practices
- 2012-19 Helen Schonenberg (TUE)
What's Next? Operational Support for Business Process Execution
- 2012-20 Ali Bahramisharif (RUN)
Covert Visual Spatial Attention, a Robust Paradigm for Brain-Computer Interfacing
- 2012-21 Roberto Cornacchia (TUD)
Querying Sparse Matrices for Information Retrieval
- 2012-22 Thijs Vis (UvT)
Intelligence, politie en veiligheidsdienst: verenigbare grootheden?
- 2012-23 Christian Muehl (UT)
Toward Affective Brain-Computer Interfaces: Exploring the Neurophysiology of Affect during Human Media Interaction
- 2012-24 Laurens van der Werff (UT)
Evaluation of Noisy Transcripts for Spoken Document Retrieval
- 2012-25 Silja Eckartz (UT)
Managing the Business Case Development in Inter-Organizational IT Projects: A Methodology and its Application
- 2012-26 Emile de Maat (UVA)
Making Sense of Legal Text
- 2012-27 Hayrettin Gurkok (UT)
Mind the Sheep! User Experience Evaluation & Brain-Computer Interface Games
- 2012-28 Nancy Pascall (UvT)
Engendering Technology Empowering Women
- 2012-29 Almer Tigelaar (UT)
Peer-to-Peer Information Retrieval
- 2012-30 Alina Pommeranz (TUD)
Designing Human-Centered Systems for Reflective Decision Making
- 2012-31 Emily Bagarukayo (RUN)
A Learning by Construction Approach for Higher Order Cognitive Skills Improvement, Building Capacity and Infrastructure
- 2012-32 Wietske Visser (TUD)
Qualitative multi-criteria preference representation and reasoning
- 2012-33 Rory Sie (OUN)
Coalitions in Cooperation Networks (COCOON)
- 2012-34 Pavol Jancura (RUN)
Evolutionary analysis in PPI networks and applications
- 2012-35 Evert Haasdijk (VU)
Never Too Old To Learn – On-line Evolution of Controllers in Swarm- and Modular Robotics
- 2012-36 Denis Ssebugwawo (RUN)
Analysis and Evaluation of Collaborative Modeling Processes
- 2012-37 Agnes Nakakawa (RUN)
A Collaboration Process for Enterprise Architecture Creation
- 2012-38 Selmar Smit (VU)
Parameter Tuning and Scientific Testing in Evolutionary Algorithms
- 2012-39 Hassan Fatemi (UT)
Risk-aware design of value and coordination networks
- 2012-40 Agus Gunawan (UvT)
Information Access for SMEs in Indonesia
- 2012-41 Sebastian Kelle (OU)
Game Design Patterns for Learning
- 2012-42 Dominique Verpoorten (OU)
Reflection Amplifiers in self-regulated Learning
- 2012-43 Withdrawn
- 2012-44 Anna Tordai (VU)
On Combining Alignment Techniques
- 2012-45 Benedikt Kratz (UvT)
A Model and Language for Business-aware Transactions
- 2012-46 Simon Carter (UVA)
Exploration and Exploitation of Multilingual Data for Statistical Machine Translation
- 2012-47 Manos Tsagkias (UVA)
Mining Social Media: Tracking Content and Predicting Behavior
- 2012-48 Jorn Bakker (TUE)
Handling Abrupt Changes in Evolving Time-series Data
- 2012-49 Michael Kaisers (UM)
Learning against Learning - Evolutionary dy-

namics of reinforcement learning algorithms in strategic interactions

2012-50 Steven van Kervel (TUD)

Ontology driven Enterprise Information Systems Engineering

2012-51 Jeroen de Jong (TUD)

Heuristics in Dynamic Scheduling; a practical framework with a case study in elevator dispatching

2013

2013-01 Viorel Milea (EUR)

News Analytics for Financial Decision Support

2013-02 Erietta Liarou (CWI)

MonetDB/DataCell: Leveraging the Column-store Database Technology for Efficient and Scalable Stream Processing

2013-03 Szymon Klarman (VU)

Reasoning with Contexts in Description Logics

2013-04 Chetan Yadati (TUD)

Coordinating autonomous planning and scheduling

2013-05 Dulce Pumareja (UT)

Groupware Requirements Evolutions Patterns

2013-06 Romulo Goncalves (CWI)

The Data Cyclotron: Juggling Data and Queries for a Data Warehouse Audience

2013-07 Giel van Lankveld (UvT)

Quantifying Individual Player Differences

2013-08 Robbert-Jan Merk (VU)

Making enemies: cognitive modeling for opponent agents in fighter pilot simulators

2013-09 Fabio Gori (RUN)

Metagenomic Data Analysis: Computational Methods and Applications

2013-10 Jeewanie Jayasinghe Arachchige (UvT)

A Unified Modeling Framework for Service Design.

2013-11 Evangelos Pournaras (TUD)

Multi-level Reconfigurable Self-organization in Overlay Services

2013-12 Marian Razavian (VU)

Knowledge-driven Migration to Services

2013-13 Mohammad Safiri (UT)

Service Tailoring: User-centric creation of integrated IT-based homecare services to support independent living of elderly

2013-14 Jafar Tanha (UVA)

Ensemble Approaches to Semi-Supervised Learning

2013-15 Daniel Hennes (UM)

Multiagent Learning - Dynamic Games and Applications

2013-16 Eric Kok (UU)

Exploring the practical benefits of argumentation in multi-agent deliberation

2013-17 Koen Kok (VU)

The PowerMatcher: Smart Coordination for the Smart Electricity Grid

2013-18 Jeroen Janssens (UvT)

Outlier Selection and One-Class Classification

2013-19 Renze Steenhuizen (TUD)

Coordinated Multi-Agent Planning and Scheduling

2013-20 Katja Hofmann (UvA)

Fast and Reliable Online Learning to Rank for Information Retrieval

2013-21 Sander Wubben (UvT)

Text-to-text generation by monolingual machine translation

2013-22 Tom Claassen (RUN)

Causal Discovery and Logic

2013-23 Patricio de Alencar Silva (UvT)

Value Activity Monitoring

2013-24 Haitham Bou Ammar (UM)

Automated Transfer in Reinforcement Learning

2013-25 Agnieszka Anna Latoszek-Berendsen (UM)

Intention-based Decision Support. A new way of representing and implementing clinical guidelines in a Decision Support System

2013-26 Alireza Zarghami (UT)

Architectural Support for Dynamic Homecare Service Provisioning

2013-27 Mohammad Huq (UT)

Inference-based Framework Managing Data Provenance

2013-28 Frans van der Sluis (UT)

When Complexity becomes Interesting: An Inquiry into the Information eXperience

2013-29 Iwan de Kok (UT)

Listening Heads

2013-30 Joyce Nakatumba (TUE)

Resource-Aware Business Process Management: Analysis and Support

2013-31 Dinh Khoa Nguyen (UvT)

Blueprint Model and Language for Engineering Cloud Applications

2013-32 Kamakshi Rajagopal (OUN)

Networking For Learning: The role of Networking in a Lifelong Learner's Professional Development

2013-33 Qi Gao (TUD)

User Modeling and Personalization in the Microblogging Sphere

2013-34 Kien Tjin-Kam-Jet (UT)

Distributed Deep Web Search

2013-35 Abdallah El Ali (UvA)

Minimal Mobile Human Computer Interaction. Promotor: Prof. dr. L. Hardman (CWI/UVA)

2013-36 Than Lam Hoang (TUE)

Pattern Mining in Data Streams

2013-37 Dirk Börner (OUN)

Ambient Learning Displays

2013-38 Eelco den Heijer (VU)

Autonomous Evolutionary Art

- 2013-39 Joop de Jong (TUD)
A Method for Enterprise Ontology based Design of Enterprise Information Systems
- 2013-40 Pim Nijssen (UM)
Monte-Carlo Tree Search for Multi-Player Games
- 2013-41 Jochem Liem (UVA)
Supporting the Conceptual Modelling of Dynamic Systems: A Knowledge Engineering Perspective on Qualitative Reasoning
- 2013-42 Léon Planken (TUD)
Algorithms for Simple Temporal Reasoning
- 2013-43 Marc Bron (UVA)
Exploration and Contextualization through Interaction and Concepts

2014

- 2014-01 Nicola Barile (UU)
Studies in Learning Monotone Models from Data
- 2014-02 Fiona Tuliayo (RUN)
Combining System Dynamics with a Domain Modeling Method
- 2014-03 Sergio Raul Duarte Torres (UT)
Information Retrieval for Children: Search Behavior and Solutions
- 2014-04 Hanna Jochmann-Mannak (UT)
Websites for children: search strategies and interface design - Three studies on children's search performance and evaluation
- 2014-05 Jurriaan van Reijssen (UU)
Knowledge Perspectives on Advancing Dynamic Capability
- 2014-06 Damian Tamburri (VU)
Supporting Networked Software Development
- 2014-07 Arya Adriansyah (TUE)
Aligning Observed and Modeled Behavior
- 2014-08 Samur Araujo (TUD)
Data Integration over Distributed and Heterogeneous Data Endpoints
- 2014-09 Philip Jackson (UvT)
Toward Human-Level Artificial Intelligence: Representation and Computation of Meaning in Natural Language
- 2014-10 Ivan Salvador Razo Zapata (VU)
Service Value Networks
- 2014-11 Janneke van der Zwaan (TUD)
An Empathic Virtual Buddy for Social Support
- 2014-12 Willem van Willigen (VU)
Look Ma, No Hands: Aspects of Autonomous Vehicle Control
- 2014-13 Arlette van Wissen (VU)
Agent-Based Support for Behavior Change: Models and Applications in Health and Safety Domains
- 2014-14 Yangyang Shi (TUD)
Language Models With Meta-information
- 2014-15 Natalya Mogles (VU)
Agent-Based Analysis and Support of Human Functioning in Complex Socio-Technical Systems: Applications in Safety and Healthcare
- 2014-16 Krystyna Milian (VU)
Supporting trial recruitment and design by automatically interpreting eligibility criteria
- 2014-17 Kathrin Dentler (VU)
Computing healthcare quality indicators automatically: Secondary Use of Patient Data and Semantic Interoperability
- 2014-18 Mattijs Ghijsen (VU)
Methods and Models for the Design and Study of Dynamic Agent Organizations
- 2014-19 Vinicius Ramos (TUE)
Adaptive Hypermedia Courses: Qualitative and Quantitative Evaluation and Tool Support
- 2014-20 Mena Habib (UT)
Named Entity Extraction and Disambiguation for Informal Text: The Missing Link
- 2014-21 Kassidy Clark (TUD)
Negotiation and Monitoring in Open Environments
- 2014-22 Marieke Peeters (UU)
Personalized Educational Games - Developing agent-supported scenario-based training
- 2014-23 Eleftherios Sidiourgos (UvA/CWI)
Space Efficient Indexes for the Big Data Era
- 2014-24 Davide Ceolin (VU)
Trusting Semi-structured Web Data
- 2014-25 Martijn Lappenschaar (RUN)
New network models for the analysis of disease interaction
- 2014-26 Tim Baarslag (TUD)
What to Bid and When to Stop
- 2014-27 Rui Jorge Almeida (EUR)
Conditional Density Models Integrating Fuzzy and Probabilistic Representations of Uncertainty
- 2014-28 Anna Chmielewicz (VU)
Decentralized k-Clique Matching
- 2014-29 Jaap Kabbedijk (UU)
Variability in Multi-Tenant Enterprise Software
- 2014-30 Peter de Cock (UvT)
Anticipating Criminal Behaviour
- 2014-31 Leo van Moergestel (UU)
Agent Technology in Agile Multiparallel Manufacturing and Product Support
- 2014-32 Naser Ayat (UvA)
On Entity Resolution in Probabilistic Data
- 2014-33 Tesfa Tegegne (RUN)
Service Discovery in eHealth
- 2014-34 Christina Manteli (VU)
The Effect of Governance in Global Software Development: Analyzing Transactive Memory Systems.
- 2014-35 Joost van Ooijen (UU)
Cognitive Agents in Virtual Worlds: A Middleware Design Approach

- 2014-36 Joos Buijs (TUE)
Flexible Evolutionary Algorithms for Mining Structured Process Models
- 2014-37 Maral Dadvar (UT)
Experts and Machines United Against Cyberbullying
- 2014-38 Danny Plass-Oude Bos (UT)
Making brain-computer interfaces better: improving usability through post-processing.
- 2014-39 Jasmina Maric (UvT)
Web Communities, Immigration, and Social Capital
- 2014-40 Walter Omona (RUN)
A Framework for Knowledge Management Using ICT in Higher Education
- 2014-41 Frederic Hogenboom (EUR)
Automated Detection of Financial Events in News Text
- 2014-42 Carsten Eijkhof (CWI/TUD)
Contextual Multidimensional Relevance Models
- 2014-43 Kevin Vlaanderen (UU)
Supporting Process Improvement using Method Increments
- 2014-44 Paulien Meesters (UvT)
Intelligent Blauw. Met als ondertitel: Intelligence-gestuurde politiezorg in gebiedsgebonden eenheden.
- 2014-45 Birgit Schmitz (OUN)
Mobile Games for Learning: A Pattern-Based Approach
- 2014-46 Ke Tao (TUD)
Social Web Data Analytics: Relevance, Redundancy, Diversity
- 2014-47 Shangsong Liang (UVA)
Fusion and Diversification in Information Retrieval
- 2015-01 Niels Netten (UvA)
Machine Learning for Relevance of Information in Crisis Response
- 2015-02 Faiza Bukhsh (UvT)
Smart auditing: Innovative Compliance Checking in Customs Controls
- 2015-03 Twan van Laarhoven (RUN)
Machine learning for network data
- 2015-04 Howard Spoelstra (OUN)
Collaborations in Open Learning Environments
- 2015-05 Christoph Bösch (UT)
Cryptographically Enforced Search Pattern Hiding
- 2015-06 Farideh Heidari (TUD)
Business Process Quality Computation - Computing Non-Functional Requirements to Improve Business Processes
- 2015-07 Maria-Hendrike Peetz (UvA)
Time-Aware Online Reputation Analysis
- 2015-08 Jie Jiang (TUD)
Organizational Compliance: An agent-based model for designing and evaluating organizational interactions
- 2015-09 Randy Klaassen (UT)
HCI Perspectives on Behavior Change Support Systems
- 2015-10 Henry Hermans (OUN)
OpenU: design of an integrated system to support lifelong learning
- 2015-11 Yongming Luo (TUE)
Designing algorithms for big graph datasets: A study of computing bisimulation and joins
- 2015-12 Julie M. Birkholz (VU)
Modi Operandi of Social Network Dynamics: The Effect of Context on Scientific Collaboration Networks
- 2015-13 Giuseppe Procaccianti (VU)
Energy-Efficient Software
- 2015-14 Bart van Straalen (UT)
A cognitive approach to modeling bad news conversations
- 2015-15 Klaas Andries de Graaf (VU)
Ontology-based Software Architecture Documentation
- 2015-16 Changyun Wei (UT)
Cognitive Coordination for Cooperative Multi-Robot Teamwork
- 2015-17 André van Cleeff (UT)
Physical and Digital Security Mechanisms: Properties, Combinations and Trade-offs
- 2015-18 Holger Pirk (CWI)
Waste Not, Want Not! - Managing Relational Data in Asymmetric Memories
- 2015-19 Bernardo Tabuenca (OUN)
Ubiquitous Technology for Lifelong Learners
- 2015-20 Lois Vanhée (UU)
Using Culture and Values to Support Flexible Coordination
- 2015-21 Sibren Fetter (OUN)
Using Peer-Support to Expand and Stabilize Online Learning
- 2015-22 Zhemina Zhu (UT)
Co-occurrence Rate Networks
- 2015-23 Luit Gazendam (VU)
Cataloguer Support in Cultural Heritage
- 2015-24 Richard Berendsen (UVA)
Finding People, Papers, and Posts: Vertical Search Algorithms and Evaluation
- 2015-25 Steven Woudenberg (UU)
Bayesian Tools for Early Disease Detection
- 2015-26 Alexander Hogenboom (EUR)
Sentiment Analysis of Text Guided by Semantics and Structure
- 2015-27 Sándor Héman (CWI)
Updating compressed column stores

2015

- 2015-28 Janet Bagorogoza (TiU)
Knowledge Management and High Performance;
The Uganda Financial Institutions Model for HPO
- 2015-29 Hendrik Baier (UM)
Monte-Carlo Tree Search Enhancements for One-
Player and Two-Player Domains
- 2015-30 Kiavash Bahreini (OU)
Real-time Multimodal Emotion Recognition in
E-Learning
- 2015-31 Yakup Koç (TUD)
On the robustness of Power Grids
- 2015-32 Jerome Gard (UL)
Corporate Venture Management in SMEs
- 2015-33 Frederik Schadd (TUD)
Ontology Mapping with Auxiliary Resources
- 2015-34 Victor de Graaf (UT)
Gesocial Recommender Systems
- 2015-35 Jungxao Xu (TUD)
Affective Body Language of Humanoid Robots:
Perception and Effects in Human Robot Interaction
- ## 2016
- 2016-01 Syed Saiden Abbas (RUN)
Recognition of Shapes by Humans and Machines
- 2016-02 Michiel Christiaan Meulendijk (UU)
Optimizing medication reviews through decision
support: prescribing a better pill to swallow
- 2016-03 Maya Sappelli (RUN)
Knowledge Work in Context: User Centered
Knowledge Worker Support
- 2016-04 Laurens Rietveld (VU)
Publishing and Consuming Linked Data
- 2016-05 Evgeny Sherkhonov (UVA)
Expanded Acyclic Queries: Containment and an
Application in Explaining Missing Answers
- 2016-06 Michel Wilson (TUD)
Robust scheduling in an uncertain environment
- 2016-07 Jeroen de Man (VU)
Measuring and modeling negative emotions for
virtual training
- 2016-08 Matje van de Camp (TiU)
A Link to the Past: Constructing Historical Social
Networks from Unstructured Data
- 2016-09 Archana Nottamkandath (VU)
Trusting Crowdsourced Information on Cultural
Artefacts
- 2016-10 George Karafotias (VUA)
Parameter Control for Evolutionary Algorithms
- 2016-11 Anne Schuth (UVA)
Search Engines that Learn from Their Users
- 2016-12 Max Knobbout (UU)
Logics for Modelling and Verifying Normative
Multi-Agent Systems
- 2016-13 Nana Baah Gyan (VU)
The Web, Speech Technologies and Rural Develop-
ment in West Africa - An ICT4D Approach
- 2016-14 Ravi Khadka (UU)
Revisiting Legacy Software System Modernization
- 2016-15 Steffen Michels (RUN)
Hybrid Probabilistic Logics - Theoretical Aspects,
Algorithms and Experiments
- 2016-16 Guangliang Li (UVA)
Socially Intelligent Autonomous Agents that Learn
from Human Reward
- 2016-17 Berend Weel (VU)
Towards Embodied Evolution of Robot Organisms
- 2016-18 Albert Meroño Peñuela
Refining Statistical Data on the Web
- 2016-19 Julia Efremova (TU/e)
Mining Social Structures from Genealogical Data
- 2016-20 Daan Odijk (UVA)
Context & Semantics in News & Web Search
- 2016-21 Alejandro Moreno Celleri (UT)
From Traditional to Interactive Playspaces: Auto-
matic Analysis of Player Behavior in the Interac-
tive Tag Playground
- 2016-22 Grace Lewis (VU)
Software Architecture Strategies for Cyber-
Foraging Systems
- 2016-23 Fei Cai (UVA)
Query Auto Completion in Information Retrieval
- 2016-24 Brend Wanders (UT)
Repurposing and Probabilistic Integration of
Data; An Iterative and data model independent
approach
- 2016-25 Julia Kiseleva (TU/e)
Using Contextual Information to Understand
Searching and Browsing Behavior
- 2016-26 Dilhan Thilakarathne (VU)
In or Out of Control: Exploring Computational
Models to Study the Role of Human Awareness
and Control in Behavioural Choices, with Ap-
plications in Aviation and Energy Management
Domains
- 2016-27 Wen Li (TUD)
Understanding Geo-spatial Information on Social
Media
- 2016-28 Mingxin Zhang (TUD)
Large-scale Agent-based Social Simulation - A
study on epidemic prediction and control
- 2016-29 Nicolas Höning (TUD)
Peak reduction in decentralised electricity systems
- Markets and prices for flexible planning
- 2016-30 Ruud Mattheij (UvT)
The Eyes Have It
- 2016-31 Mohammad Khelghati (UT)
Deep web content monitoring
- 2016-32 Elco Vriezekolk (UT)
Assessing Telecommunication Service Availability
Risks for Crisis Organisations

Protecting the safety of its citizens is the first and foremost responsibility of government. Crisis organisations assist society when safety incidents occur, and help to prevent and limit incidents. Their effective incident management requires rapid information sharing to coordinate operations and expedite decision making. Modern crisis organisations therefore depend on telecommunication services, especially since many have adopted net-centric operations. When telecommunication services are unavailable during an incident, damage will increase and people may die. In order not to be caught unprepared, these organisations must know their telecom service availability risks: they need to perform a risk assessment.

This thesis describes a method called Raster that is tailored for this domain and its challenges. Using Raster, crisis organisations can now discover, analyse and prioritise the availability risks of the telecommunication services that they use. Crisis organisations can be better prepared, helping society to be safer.

ISBN 9789036541411

9 789036 541411