Computer assisted extraction, merging and correlation of identities with Tracks Inspector

Jop Hofste
Tracks Inspector
Fox-IT
Delft, The Netherlands
jop.hofste@fox-it.com

Hans Henseler
Create-IT Applied Research
Amsterdam University of
Applied Sciences
j.henseler@hva.nl

Maurice van Keulen
Faculty of EEMCS
University of Twente
Enschede, The Netherlands
m.vankeulen@utwente.nl

ABSTRACT

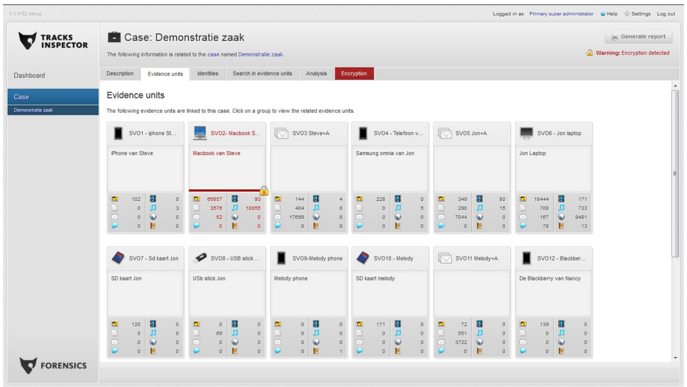
With the pervasiveness of computers and mobile devices, digital forensics becomes more important in law enforcement. Detectives increasingly depend on the scarce support of digital specialists which impedes efficiency of criminal investigations. Tracks Inspector is a commercial solution that enables non-technical investigators to easily investigate digital evidence using a web browser. We will demonstrate how Tracks Inspector can be used to discover the most important persons and groups in case data by investigators without requiring the help of digital forensics experts.

Keywords

identity extraction, identity resolution, evidence unit correlation, forensic identity research, assisted identity merging

1. INTRODUCTION

Law enforcement today relies on digital forensics in a greater variety of criminal investigations. With the pervasiveness of computers and mobile devices in society, the occurrences and volume of digital information in cases are exploding. Detectives who are intrinsically involved in collecting and assessing evidence must depend on specialists, unfamiliar with their cases, to process digital information. This impedes and even prevents prosecuting cases since there are too few digital forensics specialists and labs to support caseloads. Detectives typically investigate the evidence looking for events and information about persons. This process is essentially a review task that is similar to electronic reviews in E-Discovery projects that are described by the EDRM model [2]. Other research has revealed that technology assisted review (TAR) can greatly improve the precision and recall of relevant items [4]. Digital forensic experts acknowledge that automation and artificial intelligence can be a solution to deal with the increasing complexity and volume of digital evidence [1]. Automation is a necessary part of the solution of maintaining consistency, increasing efficiency and optimizing how digital investigators spend their time. But although these new techniques can be helpful, they also have their limitations. Ultimately, a combination of human and computer intelligence will be required. Existing TAR solutions focus on full-text search and retrieval solutions enhanced with vector-space clustering and predictive coding technologies. In contrast, this demonstration focuses on computer assisted extraction, merging and correlation of identities to assist investigators to quickly discover ”low-hanging fruit” without requiring help of a digital forensics expert.



Case: Demonstratie zaak

The following information is related to the case named Demonstratie zaak.

Warning: Encryption detected

Evidence units

The following evidence units are linked to this case. Click on a group to view the related evidence units.

SVO1 - iphone St... — iPhone van Steve — 102, 0, 0, 0 | 0, 3, 0, 0
SVO2- Macbook S... — Macbook van Steve — 66857, 3576, 52, 0 | 93, 10055, 0, 0
SVO3 Steve+A — 144, 404, 17699, 0 | 4, 0, 0, 0
SVO4 - Telefoon v... — Samsung omnia van Jon — 228, 0, 0, 0 | 0, 5, 0, 0
SVO5 Jon+A — 349, 290, 7044, 0 | 80, 15, 0, 0
SVO6 - Jon laptop — Jon Laptop — 18444, 708, 167, 78 | 171, 733, 9481, 13
SVO7 - Sd kaart Jon — SD kaart Jon — 120, 0, 0, 0 | 0, 0, 0, 0
SVO8 - USB stick ... — USb stick Jon — 0, 69, 0, 0 | 0, 0, 0, 0
SVO9-Melody phone — Melody phone — 0, 0, 0, 0 | 0, 0, 0, 1
SVO10 - Melody — SD kaart melody — 171, 0, 0, 0 | 0, 0, 0, 0
SVO11 Melody+A — 72, 551, 3722, 0 | 0, 0, 0, 0
SVO12 - Blackber... — De Blackberry van Nancy — 139, 0, 0, 0 | 0, 0, 0, 0

Figure 1: Tracks Inspector case dashboard with evidence units

The algorithms proposed here have been implemented in Tracks Inspector [5]. This is a commercial solution (figure 1) that enables non-technical investigators to easily investigate digital evidence using a web browser. Tracks Inspector brings simplicity, scalability and collaboration to the handling, storage, processing, management and reporting of digital evidence. While not intended to replace laboratory-quality solutions such as FTK and EnCase, Tracks Inspector provides a complementary solution to solve more cases and solve them faster by reducing the workloads on digital specialists to only the most complex cases.

2. TRACKS INSPECTOR

Tracks Inspector supports multiple cases that each can contain multiple evidence units. The system can receive several different input formats: disk images, directories and physical devices, as well as well-known forensic image formats such as Encase image files. The input is automatically explored and processed with open source components

ICAIL ’13, June 10-14, 2013, Rome, Italy
Copyright 2013 ACM 978-1-4503-2080-1/13/06.

in robust processes that can run on a distributed network of Linux-based servers. Meta data extracted from evidence units is stored in a MySQL database. Based on the file type, specific file meta data is extracted and file contents are converted to a HTML5 compatible format. The files are categorized into eight main categories: pictures, video, audio, documents, email, internet history, contacts and communication.

3. IDENTITY EXTRACTION

We define identity extraction as the extraction of possible identities from digital evidence. An identity is an object which is intended to refer to one single real world person. An identity representation can be generated by analyzing sources that mention real world persons. An identity is identified by its name and can be associated with related information. Currently, we assume that identity names are unique. In reality this is not always the case as people can have the same name. This is a well-known problem in, for example, co-author resolution of publications. The surnames in the languages Korean and Chinese are quite often similar and therefore it is quite difficult to determine which person is meant [7]. Since the scope of one forensic case and the impact of the problem are limited, this simplifying assumption only sporadically hinders.

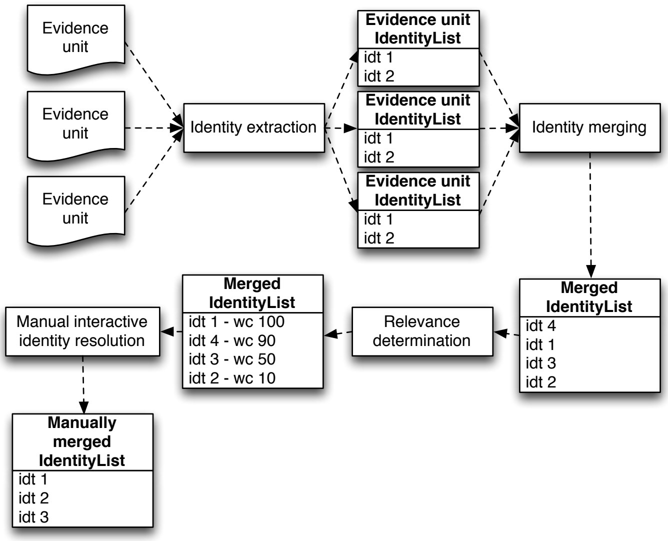


Figure 2: Simplified identity extraction and merging process in Tracks Inspector

The process starts with the extraction of identities (figure 2). The algorithm focuses on identity extraction from structured data sources, e.g., system accounts, email headers, document meta data, address books, registry settings, cookies, internet history urls and headers from chats, phone calls, text messages and other communications. This type of extraction requires access to the logical file system and knowledge about the operating system. This is different than known other approaches such as feature extraction implemented in the Bulk Extractor [3] which extracts large unique strings such as email addresses, social security numbers and credit card numbers from raw disk sectors. Each file type has its own specific type of extraction that stores extracted identities in a database. For some sources the identities are extracted in an early stage of evidence unit processing, e.g., for operating system accounts. This approach is highly scalable because the identity extraction process is part of the standard processing so that files can be analyzed in a single pass [6].

4. DEMONSTRATION

Experiments proved that identity extraction does not impact the scalability of Tracks Inspector as the overhead of identity extraction is less than 1% in the total processing of case data. Furthermore, Tracks Inspector competes well with systems like Clearwell and Trident on identity discovery, sometimes even discovering more identities and aliases such as account names because of its support for a broad range of memory and file types. Validation with a real historic forensic case also showed [6] that Track Inspector easily discovered all identities it was expected to find and that the forensic researchers on the case were impressed by the immediate insight they got into co-occurrence of identities. Investigators recommend also the way of sorting identities, various sorting options are available and these give a clear overview of which identities are important in a specific case. Identity extraction, merging and correlation in Tracks Inspector will be demonstrated using a working system with a case containing evidence that has already been processed. The demonstration will explain the basic mechanisms for processing evidence and the use of dashboards to guide investigators in their investigation. The identities dashboard and analysis dashboard will be explained in detail and an overview of experimental results as well as real case results will be presented.

5. REFERENCES

[1] E. Casey. Automation and artificial intelligence in digital forensics. *EAFS2012*, Aug. 2012. Abstract published in http://www.eafs2012.eu/sites/default/files/files/abstract_book_eafs2012.pdf.

[2] R. Doe. The e-dicsovery reference model (edrm). the review stage., Dec. 2010.

[3] S. Garfinkel. Forensic feature extraction and cross-drive analysis. *digital investigation*, 3:71–81, 2006.

[4] M. Grossman and G. Cormack. Technology-assisted review in e-discovery can be more effective and more efficient than exhaustive manual review. *Rich. JL & Tech.*, 17:11–16, 2011.

[5] J. Henseler, J. Hofste, and A. Post. Tracks inspector: Putting digital investigations in the hands of investigators. *Submitted to the ISDFS 2013*, 2013.

[6] J. Hofste. Scalable identity extraction and ranking in tracks inspector. Master's thesis, Univ. of Twente, November 2012.

[7] T. Velden, A. Haque, and C. Lagoze. Resolving author name homonymy to improve resolution of structures in co-author networks. In *Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries*, pages 241–250. ACM, 2011.