ERICSSON

Uppgjord (även faktaansvarig om annan) - *Prepared (also subject responsible if other)*		Nr - *No.*		
EMN/K/A Martin van der Zee (5407)		1/0362-FCP NB 102 88 Uen		
Dokansv/Godk - *Doc respons/Approved*	Kontr - *Checked*	Datum - *Date*	Rev	File
EMN/K/A Geert Heijenk (5430)		1999-07-08	A	

`

Quality of Service Routing

State of the Art Report

Summary

Quality of Service (QoS) routing is concerned with routing traffic such that the QoS requirements of the carried traffic are met. This report gives an overview of the different proposals for QoS-based routing. Moreover it provides a state of the art of QoS-based routing solutions in both the IETF and the literature. To set the scene, the routing mechanisms used in PSTN, ATM and IP networks are discussed first. Next the issues involved in QoS-based routing are identified, providing a framework to evaluate QoS-based routing solutions. Next QoS-based routing solutions, related to the widely deployed intra-domain OSPF routing protocol, are evaluated. As QoS-based routing solutions, in many cases, are related to the Integrated Services concept, the interface between QoS routing and RSVP is discussed in detail. Finally miscellaneous issues w.r.t. QoS-based routing are evaluated. At the end of the report conclusions are drawn and a list of patents, concerning QoS-based routing, is provided.

Contents

1 Abbreviations

AS	Autonomous System
BE	Best Effort
BGP	Border Gateway Protocol
CAC	Call Admission Control
DABRA	Deterministic Area Border Router
DNHR	Dynamic Non-Hierarchical Routing
DTL	Designated Transit List
ECMP	Equal-Cost Multipath
FEC	Forward Error Correction
IDRP	Inter-Domain Routing Protocol
IETF	Internet Engineering Task Force
IGP	Interior Gateway Protocol
IS	Integrated Services
LER	Label Edge Router
LDP	Label Distribution Protocol
LSA	Link State Advertisement
LSD	Link State Database
LSP	Label Switched Path
LSR	Label Switching Router
MF	Multi-Field
MOSPF	Multicast Open Shortest Path First
MPLS	Multi-Protocol Label Switching
NHLFE	Next Hop Label Forwarding Entry
NP	Near Polynomial
MTU	Maximum Transmission Unit
OSPF	Open Shortest Path First
PIM	Protocol Independent Multicast
PNNI	Private Network-to-Network Interface
PTSE PNNI	Topology State Element
RES-LSA	RESource Link State Advertisements
RIP	Routing Information Protocol
RRA	Resource Reservation Advertisement
RSVP	ReSource reserVation Protocol
RTT	Round Trip Time
SAR	Segmentation And Re-assembly
QOSPF	Quality of Service Open Shortest Path First
QoS	Quality of Service
QoR	Quality of Route
TCP	Transport Control Protocol
TE	Traffic Engineering
TOS	Type of Service
TTL	Time To Live
VBR	Variable Bitrate Service
VC	Virtual Channel
VP	Virtual Path

2 Introduction

Traditional Internet routing protocols calculate the shortest path based on a single metric (e.g. administrative weight or hop count) to determine the path between source and destination. This routing scheme has been successfully deployed in the last two decades for routing of Best Effort traffic in the Internet. However, the emergence of (realtime) services, requiring Quality of Service better than Best Effort, raises the question whether path selection can be improved when QoS requirements are taken into account. Furthermore, the Integrated Services concept, which provides flow based resource reservations, raises the need for routing flows. QoS-based routing can be described as:

A routing mechanism that determines the path of a flow based on knowledge of both the network resource availability and the QoS requirements of the flow.

For users the benefit of QoS-based routing is that with an increased probability a path, that supports the QoS requirements of the flow, is found. More flows are successfully established, which implies increased revenue for the network operator. Of course this benefit comes at a cost, which will be discussed extensively in the following chapters.

As normal routing is based on the shortest-path, traffic loads can be unevenly distributed in the network. Traffic engineering is concerned with provisioning sufficient resources to support these traffic requirements. Furthermore, pre-configured link metrics, such as administrative weights, can be manipulated such that some control is given over how traffic flows through the network. However, QoS-based routing provides this function automatically i.e. assigns traffic to links where there are sufficient resources to meet the QoS requirements.

One important question with QoS-based routing remains: Do we really need it? One may argue that bandwidth is abundant and proper network (over-)provisioning can eliminate potential bottlenecks, so there is no need for this extra complexity. On the other hand, one may argue that bandwidth is not abundant yet and never will be. And although the long term network capacity may be sufficient, short term periods of congestion can arise. Furthermore, link failures may destroy the perfect match between supply and demand. Therefore there is a need for QoS-based routing. Whether QoS-based routing is the right answer and whether the gain of QoS-based routing exceeds the cost, remains to be seen.

In the following text it is assumed that the reader is familiar with the TCP/IP protocol suite, IP routing, the Integrated Services and Differentiated Services concepts. References are provided for further reading in the appropriate sections.

First, an overview of existing routing protocols is provided in Chapter 3. Next the QoS-based routing proposals and issues related to QoS-based routing are evaluated in Chapter 0. The Multi-Protocol Label Switching (MPLS) framework provides some QoS-based routing capabilities, referred to as traffic engineering, which are explained in Chapter 5. Finally some conclusions concerning QoS-based routing are drawn in Chapter 6. A list of patents related to QoS-based routing is provided in Appendix A.

It is explicitly noted that the solutions presented in this document are not the ideas of the author, but merely an overview of existing work is presented, appropriate references are provided.

3 Existing routing protocols

A survey of routing protocols, used in PSTN, ATM and IP networks is provided in [IDROUTE98]. This survey gives an overview of the background from which QoS-based routing has evolved. This survey is evaluated briefly in this chapter. The survey has been extended with an evaluation of the OSPF routing protocol in IP networks. First some general definitions of network routing are provided in Section 3.1. Next the PSTN, ATM and IP routing protocols are discussed in the subsequent sections.

3.1 Definitions

Routing schemes can be divided into *static* and *dynamic* schemes. With *static* routing the routing tables are fixed for a relatively long period. Usually these routing tables are manually configured, and packets from source to destination follow the same route. This scheme does not adapt to network changes automatically which is a major drawback. On the other hand *dynamic* routing schemes, also referred to as *adaptive* routing schemes, adapt to network changes. For this purpose network connectivity information, is disseminated through the network to populate the forwarding tables of the routers. When a link failure is detected, this information is disseminated and packets are re-routed around the failed link. This feature comes with a cost in terms of protocol complexity, transmission and processing overhead and potential risks of oscillations, looping, etc [STALL98].

Two important types of dynamic routing schemes are: *distance-vector* (e.g. RIP) and *link-state* (e.g. OSPF) routing schemes. With the *distance-vector* routing scheme, each node exchanges with its neighbouring nodes its 'distance' (e.g. hop count) to other networks. The neighbouring nodes use this information to determine their 'distance' to these networks. Subsequently these nodes share this information with their neighbours, etc. In this way the reachability information is disseminated through the network. Eventually each node learns, which neighbour (i.e. next hop router) to use, to reach a particular destination with a minimum number of hops. A node does not learn about the intermediate nodes to the destination[1]. With *link-state* routing each node builds a complete topology database of the network. This topology database is used to calculate the shortest path. Each node in the network transmits its connectivity information to each other node in the network. This type of information exchange is referred to as flooding. This way each node is able to build a complete topological map of the network.

Routing schemes can also be divided into *source* routing and *hop-by-hop* routing schemes. With *source* routing there is a single node, i.e. the source node, that makes the routing decisions i.e. determines the end-to-end path. The end-to-end routing information is added to the packet to be transmitted from source to destination. The intermediate routers examine the routing information contained in the packet and forward the packet accordingly. With *hop-by-hop* routing each router determines the next-hop of the packet it receives. Therefore each router is involved in making routing decisions and each router needs to have the topology information to determine the next-hop.

The source routing information can be appended to an IPv4 packet in the options field. Either strict-source routing or loose-source routing can be selected. The strict-source route specifies the exact path (routers not specified may not be traversed), the loose-source route specifies a set of intermediate routers that must be traversed (routers not specified may be traversed). Both options are limited by the option field length, limiting the number of intermediate routers to nine. Furthermore, when IP options are used, a serious performance penalty concerning IP forwarding performance should be taken into account, as routers are optimised to forward packet with no options. Routers applying address filtering

[1] Therefore source routing cannot be supported by distance-vector routing.

for security reasons may drop packets with source routing information, as the destination address is buried in the IP options field.

The path computation involved with routing can be divided into single path and multipath route computation. With single path computation only one path is computed between source and destination. With multipath computation several paths are computed between source and destination (in case they exist), and one path is selected from this set. Multipath routing allows the traffic load to be divided over different paths i.e. apply load balancing. To preserve the ordering of packets sent by a single TCP connection with multipath routing, a router generally applies some hash function[2] to the TCP connection identifier (source and destination IP address) to determine the next hop [MOY98].

3.2 PSTN routing

The PSTN routing method is based on the E.164 numbering scheme. An E.164 number is used for routing towards the destination end-point. Four different types of routing schemes are identified: fixed routing, time-dependent routing, state-dependent routing and event-dependent routing.

With fixed routing the route selection is determined on a preplanned basis and maintained over a long period of time. Fixed routing is applied efficiently when the network is non-hierarchical i.e. flat.

With time-dependent routing the routing tables are altered at predetermined times during a day or week. Time-dependent routing adapts to the periodic changes of traffic loads. The adaptations are usually pre-configured. An example is the Dynamic Non-Hierarchical Routing (DNHR) scheme where alternate path are statically configured, subject to time-of-day changes. Time-dependent routing is an example of static routing, as the routing tables are fixed.

With state-dependent routing the routing tables are altered automatically according to the state of the network. Network state information is collected at a central processor or distributed to switches in the network. Network state information may be updated periodically or on-demand. The state-dependent routing scheme routes connections on the best available path.

With event-dependent routing the routing table is updated locally dependent on the result of routing decisions. When a connection is routed unsuccessfully on the current shortest path, an alternative path is selected. Both state- and event-dependent routing are examples of dynamic routing as the routing tables change dynamically.

In the PSTN network resource management and routing may be implicitly connected i.e. when a path is found resources are reserved otherwise the connection is blocked. For further details concerning PSTN bandwidth management see [ASH98].

3.3 ATM routing

The Private Network-to-Network Interface (PNNI) routing protocol [AF96c] is used in ATM networks to route connections based on the network state and topology information. It is the only standardised QoS-based routing protocol. PNNI is a link-state routing protocol, which uses source routing.

Each switch maintains a neighbouring relation with its neighbours through the exchange of HELLO messages. The switch/link state information and reachability information is flooded in PNNI Topology State Elements (PTSE), which are reliably transmitted within the peer group. PTSEs are transmitted

---

[2] The hash function provides the ability to balance the load over several outgoing links, while each TCP connection always selects the same outgoing link.

periodically or when certain events occur e.g. significant change in network state. The rate with at which state information is flooded can be controlled by properly setting thresholds.

A QoS path is established during the connection establishment phase. The source switch determines the end-to-end path based on the QoS requirements and the network state information. The source switch includes the selected path in the Designated Transit List (DTL) of the Setup message. Each intermediate switch performs Connection Admission Control (CAC) to incoming setup requests. CAC evaluates the QoS requirements of the setup message and the current network traffic parameters. In case CAC fails due to congestion, the path is crankbacked to the previous switch from which an alternate path is selected i.e. source routing is overruled. In [CIDON97] it is suggested to compute multiple paths and reserve resources along different paths to speed the connection establishment (no crankback).

The PNNI routing mechanism is scalable to large networks through hierarchical routing. To explain the principle of hierarchical routing, first some definitions are given [GUERIN97], [ORDA99]. The network is represented by a graph G(V, E), where V is the set of layer 1 nodes (e.g. node 1.1.1) and E is the set of layer 1 links interconnecting these nodes. G(V, E) is also referred to as the actual network. Layer 1 nodes are clustered to form layer 2 node (e.g. node 1.2), which are clustered into layer 3 nodes (e.g. 1), etc, up to the last layer. Nodes at the layer i (e.g. node 1.1.1, 1.1.2 and 1.1.3) that are clustered into the same layer i+1 node, are denoted as peers.

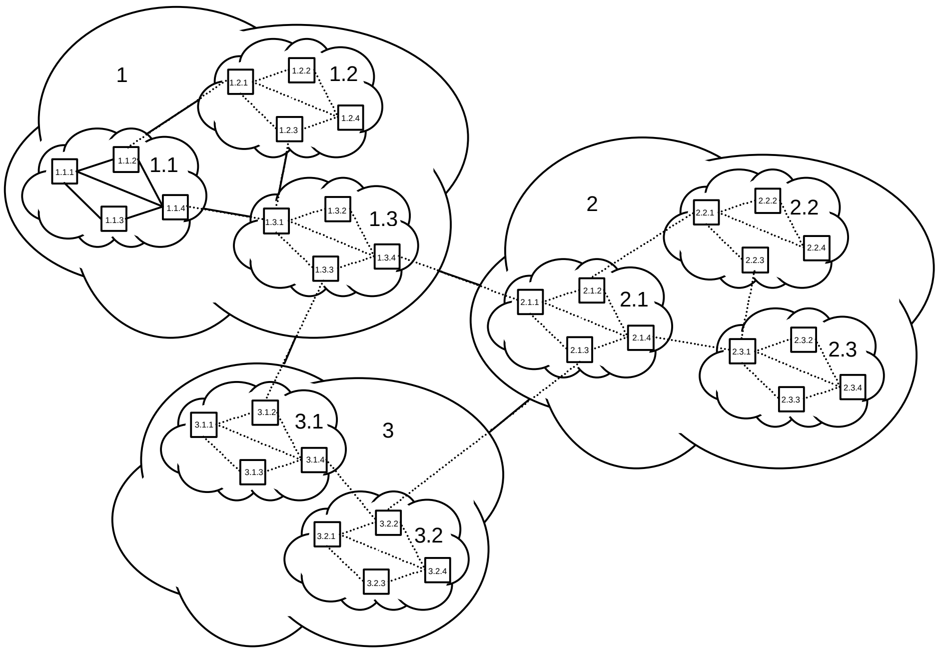

Figure 1 Hierarchical routing example.

The hierarchical routing is explained through an example depicted in Figure 1. With hierarchical routing each layer 1 node i keeps an aggregated representation of the network, referred to as the image at node i, instead of a full network state database with non-hierarchical routing. The network state information stored at node i becomes progressively aggregated as the network becomes more remote. This is exemplified in Figure 1, where the connectivity information stored by the peer nodes of cluster 1.1 is depicted through solid lines. The connectivity information indicated by the dashed lines is hidden to these peer nodes.

Each border node (e.g. node 1.1.2) at a peer group hides the details of the group to the next hierarchical level and handles inter-level interaction at call setup. The information reduction is also referred to as abstraction. Topology and state information is flooded within the peer group. The peer group border switch abstracts the information from within the group and floods the abstracted topology information to the next higher level in the hierarchy including aggregated reachable address information. This implies that in case of hierarchical routing, QoS path computation is based on aggregated i.e. inaccurate state information. The effect of abstraction on path computation is evaluated in [GUERIN97]. For example the amount of bandwidth advertised by a remote network, or set of networks, implies only that a connection request for this amount of bandwidth is *likely* to succeed. Bandwidths of individual links cannot be advertised and the advertised bandwidth may for example refer to the average bandwidth available. It is noted that the inaccuracy increases with the aggregation level i.e. scalability comes with a price.

When the network is flat i.e. if there is a single peer group, the originating switch controls the end-to-end path in the ATM network. When there are multiple levels of hierarchy, each border switch controls the part of the end-to-end path that is routed on that level of hierarchy i.e. within that peer group.

3.4 IP routing

For routing purposes the Internet is partitioned into a disjoint set of Autonomous Systems (AS) as depicted in Figure 2. Each Autonomous Systems is administered by a single authority and employs a single routing protocol.

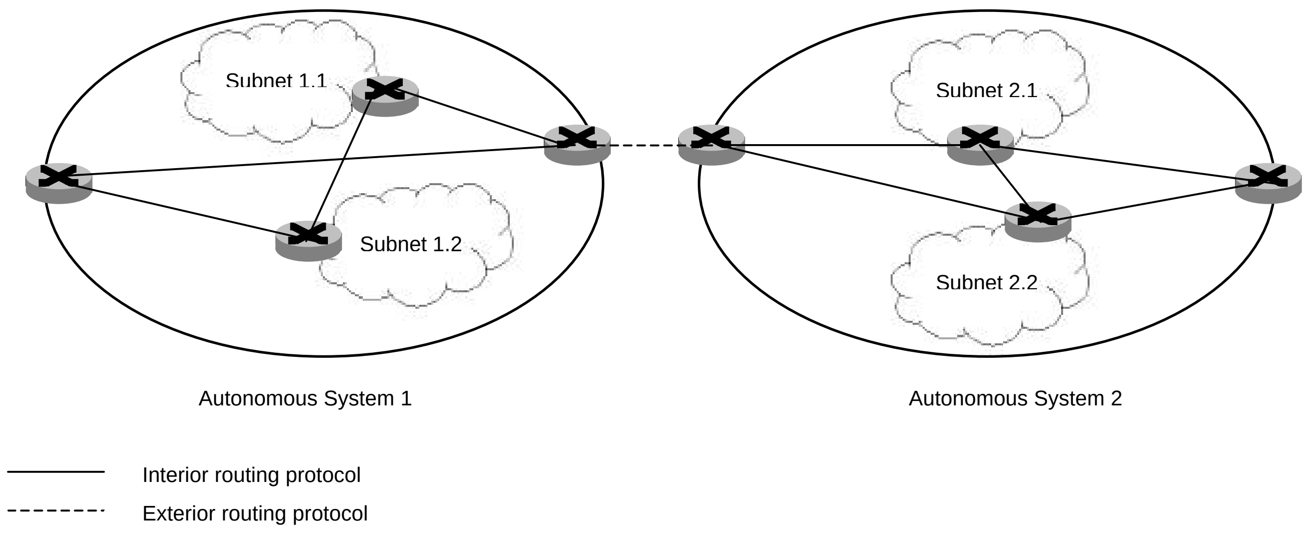

Figure 2 Autonomous Systems.

The Internet provides a connectionless data service by which datagrams are generally routed hop-by-hop. Routing is destination-based or more precisely on the network prefix contained in the destination address. The network prefix of incoming datagrams, is compared with entries in the routing table to find the next-hop router. This routing table lookup generally applies a longest-match search criteria. Although IPv4 does support source routing, many router implementations suffer from a performance penalty when source routing is used. This is an important reason why source routing is not widely deployed.

The current Internet routing behavior has been extensively analysed by means of 40,000 end-to-end route measurements between 37 Internet sites using 'traceroute' [PAXSON97]. These measurements have shown that about 2/3 of the Internet paths are stable for either days or weeks. Furthermore, loops have been detected when the network is in a transient state i.e. while routing information is disseminated, triggered by a network connectivity change. Over three days 60 instances of looping have been detected some lasting several hour and a few several days. Routing loops add to the router congestion. The measurements also revealed that around half of the routing paths where asymmetrical i.e. the path from A to B was at least one node different from the path from B to A.

Intra-domain routing employs an interior routing protocol, such as Open Shortest Path First (OSPF) and Routing Information Protocol (RIP), and inter-domain routing requires a exterior routing protocol such as Border Gateway Protocol (BGP) and Inter-Domain Routing Protocol (IDRP).

The inter-domain routing protocols BGP and IDRP both are path-vector protocols. The routing information exchanged between ASs includes vectors identifying the intermediate ASs through which a certain destination can be reached. No link metrics are exchanged, as each AS may have a different interpretation of a metric. Only reachability information is disseminated. Inter-domain routing protocols need to scale to very large networks and usually employ hierarchical routing to satisfy this requirement.

RIP is a distance-vector routing protocol which is simple and suitable for small Internets. Each router communicates with its neighbours to learn about the network connectivity. Each router periodically updates its neighbours about the cost of routes its knows to certain destinations i.e. exchanges distances. In this way the network connectivity information is propagated through the network. Limitations of the RIP protocol are that a single metric is used and the hop distance is limited (in practice to15) making it not suitable for large Internets. Furthermore, the protocol converges slowly after link failure.

Due to the limitations of RIP, OSPF is now the recommended routing protocol for intra-domain routing. OSPF is a link-state routing protocol that can be applied efficiently in large Internets. OSPF provides hop-by-hop routing, fast dissemination of routing information, multiple path calculation and hierarchical routing. OSPF applies the distributed map concept: each node has a copy of the network topology database which is updated regularly. OSPF has good convergence properties: when the network changes new routes are quickly found with a minimum of routing protocol overhead. For this purpose network connectivity information, contained in so-called Link State Advertisements (LSA), is periodically flooded in the network i.e. each router duplicates the incoming information on each of its outgoing interfaces (except the one it arrived on). In this way each router quickly builds a Link State Database (LSD).

OSPF uses the well-know Dijkstra algorithm to calculate the shortest path. When there are multiple equal cost paths between source and destination, the Dijkstra algorithm determines these paths i.e. Equal-Cost MultiPath (ECMP) [RFC2328]. This allows alternate routing of traffic also referred to as load balancing. Up to five different metrics can be selected, corresponding to the TOS field of IPv4, as listed in Table 1.

The TOS capabilities of a router are advertised in the OSPF Hello messages, exchanged between neighbouring routers. The T-bit in the OSPF options field indicates whether the router supports TOS zero only or the TOS routing options listed in Table 1. When a non-zero TOS shortest path is computed, routers with zero TOS only are skipped, i.e. these routers will not be part of the path.

Metric	**Meaning**	**TOS field**
Normal (default)	administrative weight or hop-count (pre-configured)	0
Minimise monetary cost	link cost (pre-configured)	2
Maximise reliability	reliability (pre-configured, dynamic)	4
Maximise throughput	bit duration (pre-configured)	8
Minimise delay	propagation + queueing delay (dynamic)	16

Table 1 OSPF metrics.

OSPF routing provides a simple form of QoS-based routing by routing on the Type of Service (TOS) field of the IP packet. A different shortest-path tree is computed for each of the five TOS values. OSPF allows classes of traffic to follow different routes. However, overbooking may occur and thus QoS guarantees are not provided. This in contrast to the QoS-based routing solutions, where routing takes into account the QoS requirements of the flow. It is noted that TOS routing has never been widely deployed in the Internet. Therefore TOS routing has been removed from the latest OSPF specification (version 2) [RFC2328] and is not included in the IP version 6 specification [RFC1883]. Furthermore, the Differentiated Services architecture has redefined the TOS field in IPv4 as the Differentiated Service (DS) field. The Differentiated Services architecture assumes that the TOS marking, as defined in RFC1349, is deprecated.

In [IDROUTE98] it is assumed that Multi-Protocol Label Switching (MPLS), see Section 5.1, will be deployed in the future IP networks. The interworking between PSTN/ATM routing and MPLS routing is more natural, than interworking with OSPF, because both are connection-oriented in nature.

4 QoS-based routing in IP networks

4.1 Introduction

Some people see QoS-based routing as the missing piece in the evolution of QoS-based service offering in the Internet [RFC2386]. Many of the QoS-based routing proposal are tied to the Integrated Service concept i.e. establish a QoS-path for a flow. In this chapter an overview of the different QoS-based routing proposals is provided.

First the Integrated Services concept is explained in Section 4.2. Next in a framework, presented in Section 4.3, the issues and requirements of QoS-based routing are evaluated. The intra-domain QoS-based routing proposals, based on the OSPF protocol, are evaluated in Section 4.4. The interface between QoS-based routing and the resource reservation protocol RSVP is discussed in Section 4.5. Finally some miscellaneous issues are evaluated in Section 4.6.

4.2 Integrated Services

The Integrated Service (IS) architecture [RFC1633], [WHITE97] enhances the traditional Internet architecture to provide a Quality of Service better than Best Effort. The Integrated Service architecture aims to support real-time services such as voice and video. These services cannot be supported adequately by the Best Effort service. For this purpose resources are reserved within the network for a flow, which is to receive an enhanced Quality of Service. Resources are reserved through a resource reservation protocol such as ReSource reserVation Protocol (RSVP). Through the exchange of RSVP messages between source and receiver(s), so-called reservation state is installed in the network i.e. at intermediate nodes. This reservation state is maintained per flow i.e. per data stream flowing between source and destination. Resources are reserved for each direction of the flow independently. Furthermore, both unicast and multicast transmission are supported. This is exemplified in Figure 3.

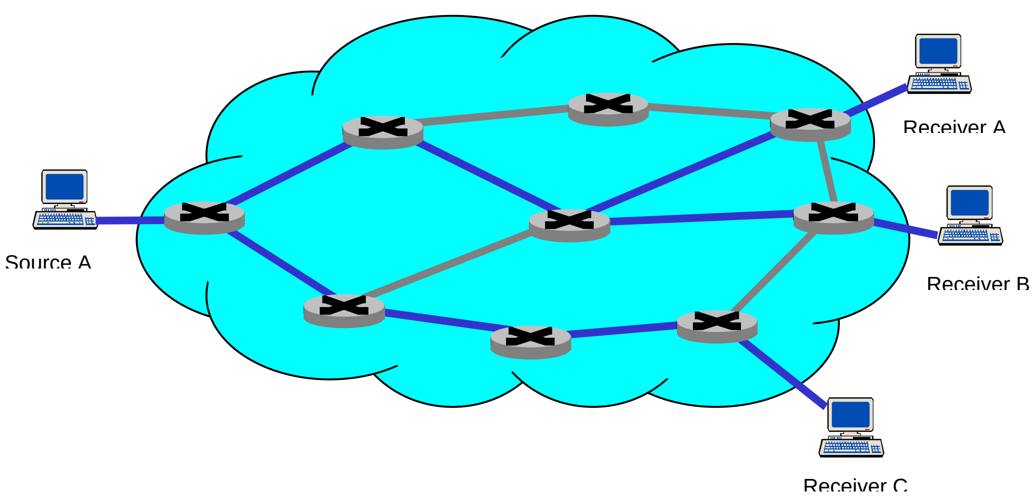

Figure 3 Resource reservation for a multi-cast flow.

The resource reservation is initiated by the periodic transmission of so-called Path messages by the source. This path message is transmitted to all receivers. The Path message is routed as any normal IP packet from source to destination. In this example there are multiple receivers, therefore multicast routing is applied. The source identifies itself to the receivers with this Path message. The Path message contains a specification of the traffic characteristics of the source, and while travelling from source to destination, collects information about the end-to-end path, which is of interest to the receiver. The Path message does not contain the actual resource reservation request itself. Note that Path messages are duplicated at branch points in the distribution tree[3].

Upon receipt of the Path message by the receiver, the receiver may decide to reserve resources for this traffic flow based on the information contained in the Path message. The receiver starts transmitting periodically so-called Resv messages towards the source. This Resv message contains the request to reserve resource along the path towards the source. It is important that the Resv message follows the same path as the Path message did, and because routing in the Internet is not necessarily symmetric, normal routing cannot be applied to route the Resv message towards the source. To facilitate this requirement, the Path message has installed so-called reserve path routing state in the routers while it travelled toward the receiver. This information is used to route Resv messages back towards the source.

While the Resv message is transmitted from receiver to the source, resources are reserved at the intermediate routers. Both the Path and the Resv messages need to be transmitted periodically (default every 30 seconds) to maintain the reservation state in the network. The network reservation is installed through so-called soft-state i.e. when the transmission of either Path or Resv messages is stopped, the resource reservation is automatically deleted[4]. This soft-state provides some of the traditional robustness of the Internet.

When a link fails along the path of the flow, the normal routing protocol will route the Path messages along a new path. This will install new reserve routing state for Resv message and eventually a new resource reservation will be made along the new path. This transition is unlikely to be seamless to the users of the service.

The Integrated Services concept makes use of the existing routing functionality in the network for routing Path messages. Thus an RSVP flow follows the same route as normal Best Effort traffic would. When no sufficient resources are found along the selected path, RSVP session establishment fails.

The Integrated Services model supports receiver heterogeneity i.e. receivers with different QoS requirements. The branches of the multicast distribution tree may have different amounts of reserved bandwidth. In this way for example receivers with both high and low QoS requirements can be supported efficiently. Furthermore, different reservation styles can be applied which refer to the way in which reservations request of receivers are merged in so-called merge points of the distribution tree.

4.3 Framework

The issues and requirements related to QoS-based routing are discussed and a framework is presented in [RFC2386]. An evaluation of QoS-based routing can also be found in [SZABO98].

---

[3] The Adspec contained in the Path message can be different, because it depents on the outgoing link characteristics.

[4] Pathtear and Resvtear messages may be used to explicitly delete the resource reservation (compared to the implicit soft-state release).

First the QoS-based routing objectives and issues with QoS-based routing in general are evaluated. Next issues specific for intra-domain routing, inter-domain routing and multicast routing are evaluated. Finally some points to consider when comparing QoS-based routing schemes are presented.

The following QoS-based routing objectives are listed [RFC2386]:

1. Path selection should be based on both the network conditions (e.g. resource availability) and QoS requirements of the flow.

2. Optimisation of network resource usage.

3. Graceful performance degradation under overload conditions.

It is noted that selecting the path with the best QoS may potentially deteriorate the overall network performance as paths are routed over long paths occupying relatively large amount of network resources.

The following issues need to be considered with QoS-based routing:

Granularity of routing

The granularity of routing refers to the information upon which routing is performed. Traditional Internet routing is based on the destination address. However, many multicast routing protocols require the combination of source and destination address (e.g. Multi-cast Open Shortest Path). The Integrated Services concept is based on per flow resource reservation (unicast and multi-cast destination address based).

The granularity of routing has an impact on the amount of routing state, i.e. routing table size, that has to be maintained at the routers. A large routing table increases the routing lookup delay. To support the Integrated Services concept, routing based on flow identification provides the optimum granularity.

QoS-metrics

The QoS-metrics of certain links can be difficult to determine, e.g. broadcast multi-access links or ATM Virtual Connections. The PNNI routing mechanism, which provides QoS-based routing in ATM, could possibly provide the required QoS metrics for ATM [AF96c].

QoS-metrics should be selected carefully to map to the QoS requirements of the flow.

The Integrated Service architecture identifies currently two classes of service (Guaranteed Service and Controlled Load Service). Each service class can experience a different QoS-metric on the same link (e.g. delay) due to service dependent handling of the routers.

QoS-metrics which are considered useful are: bandwidth[5], delay and jitter. QoS-metrics apply to both the node and the outgoing link of that node. A uniform representation across different routing domains is required.

---

[5] There are two basic option to define bandwidth of a link: bandwidth not reserved for flows or the actual link utilisation (including all traffic). The first option only applies when there is a resource reservation protocol.

Metrics can be divided into three classes [XIAO99]. Let *d(i,j)* be a metric for link *(i, j).* For any path *P = (i, j, ..., l, m),* metric *d* is:

Additive if $d(P) = d(i, j) + d(j, k) + ... + d(l, m)$

Multiplicative if $d(P) = d(i, j) * d(j, k) * ... * d(l, m)$

Concave if $d(P) = min\{d(i, j), d(j, k), ..., d(l, m)\}$

According to this definition, the metrics delay, jitter, cost and hop count are additive. The metric reliability is multiplicative and the metric bandwidth is concave.

In [XIAO99] it is argued that bandwidth and hop count are more useful constraints compared to delay and jitter as many application tolerate occasional violations of the latter constraints. Furthermore, delay can be determined when the bandwidth and traffic characteristics are known (see fluid flow model [RFC2212] which is used to determine the delay bound for the Guaranteed Service).

Optimal path computation on two or more independent QoS-metrics is in general computationally intractable i.e. NP-complete [GAREY79]. Computational complexity with more than one parameter can be reduced by utilising 'sequential filtering'. Under this approach, a combination of metrics is ordered in some fashion, reflecting the importance of the different metrics (e.g. bandwidth followed by delay, etc). Paths based on the primary metric are computed first and a subset of them are eliminated based on the secondary metric and so forth until a single path is found[6]. This is an approximation technique which trades off global optimality for path computation simplicity. Furthermore, in [MA97], [MA98] it is pointed out that the bandwidth, delay and jitter QoS-metrics are not independent but correlated in networks with rate dependent scheduling (e.g. Weighted Fair Queueing). These scheduling techniques ensure that flows get their proportional share of the link capacity. Given a path *P* with *n* hops, the traffic characteristics of the source (token bucket specification <*b, r*> with *b* bucket depth and *r* token rate) and the reserved bandwidth *R*, the delay *D*, jitter *J* and buffer space *B* at the *h*-th hop can be derived [MA98]:

$$D(P,R,b) = \frac{b}{R} + \frac{nM}{R} + \sum_{i=1}^{n}\left(\frac{M}{C_i} + l_i\right)$$

$$J(P,R,b) = \frac{b}{R} + \frac{nM}{R}$$

$$B(P,R,b) = b + hM$$

Where C_i is the capacity of link *i*, *M* the maximum packet size and l_i the latency[7] of link *i*. This dependency can be used to develop a modified Bellman-Ford algorithm that selects paths satisfying these multiple constraints. For further details see [MA97].

[6] A simpler filtering technique can be applied in case of two metrics. First those links are removed from the computation which do not satisfy one QoS requirement (e.g. bandwidth) and next calculate the shortest path for the remaining links based on the second metric (e.g. cost).
[7] As the line speeds are increasing, this latency is mainly determined by the router induced delay in forwarding the packet and to a lesser extend to the line transmission delay.

Suitable triggers, which separate a minor change from a major change in QoS-metric, need to be defined. A major change is assumed to trigger a routing update. The stability of the QoS routes is dependent on the routing update frequency which should preferably be controllable. The routing update frequency can be controlled by specifying a minimum time between updates, applying some kind of filtering to the measured QoS-metric or using appropriate thresholds (relative or absolute threshold level).

Network performance

QoS path selection should be based on both the QoS requirements of the flow and network performance objectives. As QoS paths may travel along longer routes so that, more network resources are consumed compared to traditional shortest-path routing. To ensure network efficiency a 'high level admission control, which controls the path selection from a network performance point of view, is introduced. This high level admission control should ensure that admission of one flow does not excessively reduce the probability of successful admission of future flows.

When QoS-based routing increases the average path length compared to normal routing, then there is a penalty to be paid for QoS-based routing in terms of network performance. The alternate routing scheme, used for experiments in PSTN networks, is an example of a scheme where the path length increases when the network is heavily loaded. Under high load conditions, alternate traffic along a long route competes with minimum hop traffic for link resources. Enhanced alternate routing schemes limit the path length under high load conditions e.g. by exclusively reserving link resources for direct routed calls. Another strategy of the high level admission control could be to limit the difference between the minimum path length and the QoS path length.

Administrative controls

Administrative control issues are related to distinguishing classes of service for which routing is performed. One approach is that different priority classes are identified each with a different probability for a QoS path request to be successful. Classes with higher priority could be provided a higher success probability compared to lower priority flows. Network resources are divided such that different amounts of network resources are provisioned to each class of service. Thus each class of service views a different network resource availability.

Furthermore, it is noted that in case of no priorities, flows with a low QoS requirement are more likely to succeed than flows with high QoS requirements.

QoS-based routing for multicast flows

In some points of the multicast distribution tree, resource reservations of Integrated Services are merged dependent on the reservation style. This merging process results in a distribution tree with different QoS requirements per link. Furthermore, receiver heterogeneity i.e. receivers with different QoS requirements, results in branches with different QoS requirements. An example of a multi-cast application with receiver heterogeneity is a video transmission with two levels of encoding i.e. high priority and low priority frames. Receivers who wish to receive high video quality, receive both high and low priority frames. Receivers who tolerate a lower video quality, only receive the high priority frames. Multicast QoS-based routing needs to take into account branches with different QoS requirements.

QoS-based routing overhead

Three different types of routing overhead can be identified: computation, storage and communication. The computational overhead stems from the calculation of the QoS-based paths. The storage overhead refers to the amount of data that is needed for route calculation and that has to be stored. The communication overhead includes the transmission of the network state information in the network (e.g. flooding in case of OSPF).

Scalability

A common technique to provide a scalable routing technique is hierarchical routing (See PNNI routing protocol in Section 3.3). Scalability is obtained at the expense of the accuracy of aggregated state information. A way to compensate for inaccurate information during setup is to use crankback i.e. recursively backtrack along the partial flow path up to the first node that can determine an alternative path to the destination.

When QoS routing solutions for the intra- and inter-domain are compared, the intra-domain solution could be more complex and feature rich compared to the inter-domain solution where scalability, simplicity, consistency and stability are the most important design considerations. The development of a single routing solution for both domains is not recommended due to these different requirements.

A routing protocol can aid to distribute information concerning the expected cost of QoS paths. This information can be used to charge users, that require QoS services.

QoS route selection should be based on multiple path calculation. Alternate path calculation is important, as the shortest path may not provide the desired QoS.

Route oscillation is the unwanted effect when traffic is routed alternating on two different routes. Path selection influences the metrics of the path, which subsequently triggers changes to the path selection. Thus there is a feedback loop between path selection and the link's QoS metrics. In the early ARPAnet, where routing was performed based on the measured Round Trip Time (RTT), an inability to converge to a stable state has been observed. It is noted that when the path of a route changes frequently, this can result in increased delay and jitter.

Intra-domain QoS-based routing

The following requirements apply for intra-domain QoS based routing:

- The routing scheme must route the flow along a path that accommodates its QoS requirements, or indicate that no path with requested QoS is available.
- The routing scheme must indicate disruptions to the current route in case of topological changes.
- The routing scheme must accommodate Best Effort traffic routing without any changes.
- The support of QoS-based multicasting with receiver heterogeneity is considered optional.
- The routing scheme should be capable to optimise the network resource usage.
- Implementation of high level admission control to limit the resource utilisation of individual flows.

<u>Inter-domain QoS-based routing</u>

Scalability prevents the use of dynamic network state information for inter-domain routing. The inter-domain routing solution is rather based on network engineering principles and the exchange of relatively sparse information.

The following inter-domain functions must be provided:

- Exchange of reachability information to various destinations with address aggregation whenever possible.
- Loop free routing
- Determination of the QoS on the path towards the destination. The QoS of an end-to-end path includes the QoS available in each transit AS. The QoS dynamics within the transit AS is not exchanged, but rather static QoS information is provided. The QoS metrics advertised are based on contracts for service provisioning. The QoS metric of the AS can be viewed as the QoS capabilities of a 'virtual path'. This approach aligns with the Differentiated Services concept [RFC2475].
- To reduce the flow state processing in transit ASs, flows are aggregated i.e. classified into a few aggregate service classes. This implies that traffic through the transit AS is not routed at the individual flow level, but rather based on a service class identification.
- Multiple path calculation for different service classes.
- Application of routing policies.

<u>QoS-based multicast routing</u>

The following objectives apply for multicast routing:

- Scalability to large groups with dynamic membership
- Robustness to topological changes
- Support for receiver initiated heterogeneous reservations
- Support for shared reservation styles
- Support for administrative control of the resource consumption of the multicast flow.

When considering the scalability of multicast QoS-based routing, the overhead associated with receiver discovery and the overhead associated with QoS-based path computation need to be evaluated.

The QOSPF [IDOSPF97] makes use of source routing to compute the multicast QoS path. Receiver discovery is done by receiver location broadcasts. In case source routing is applied, the multicast tree is computed based on the sender traffic advertisement. With source routing it is difficult to support receiver heterogeneity and shared resource reservation styles.

In case receiver oriented multicast routing is applied, sender traffic advertisements are transmitted to all receivers over a common BE multicast tree[8]. The receiver discovery overhead is minimised by using a scalable BE multicast routing scheme (e.g. PIM). Each receiver computes independently a unicast QoS path between source and receiver i.e. multicast path calculation is divided over n unicast path calculations. Finally routers merge the unicast reservation messages from the receivers appropriately. This receiver oriented multicast routing scheme accommodates receiver heterogeneity and shared reservation styles easily. The disadvantage is the incompatibility with the RSVP signalling model (RSVP uses reserve path state information to route Resv message from receivers to the source).

<u>The following aspects are relevant when comparing QoS-based routing schemes:</u>

- What state information is used
- What triggers the propagation of state information
- Route computation distributed or central i.e. whether source routing or hop-by-hop routing is used
- Route computation on demand, pre-computed or in a hybrid manner
- What optimisation criteria are used in path computation
- The use of alternate path calculation

Evaluation of QoS-based routing schemes should consider performance criteria such as route blocking ratio, routed bandwidth ratio or average path length. Furthermore, appropriate network topology models and traffic models should be selected carefully.

4.4 QOSPF

The initial extension to the OSPF routing protocol to support QoS-based routing is discussed in Section 4.4.1. A more detailed proposal, whose applicability has been proven through an implementation, is described in Section 4.4.2. Then a limited form of QoS-based routing through the use of so-called Equal Cost Multi-Paths (ECMP), is presented in Section 4.4.2. Finally an overlay architecture of OSPF is evaluated in Section 4.4.4.

4.4.1 Initial proposal

In [IDOSPF97] extensions to both OSPF [RFC2328] and MOSPF [RFC1584] are described to provide QoS-based routing. Furthermore, the use of a resource reservation protocol, such as RSVP, is discussed.

To support QOSPF routers not only advertise topology information but also network resource information. The network resource information includes both router and link resources. The available and reserved resource information is advertised in RESource Link State Advertisements (RES-LSAs) and Resource Reservation Advertisements (RRAs) respectively.

A Resource Reservation Advertisement (RRA) describes a router's reservations for a particular flow. The purpose of the RRA is to indicate the resources reserved by the flow to other routers such that this information is used to calculate or recalculate the distribution tree for a flow. An RRA is originated

[8] This is a multicast tree which has been established without taking into account QoS requirements or link state information.

whenever the router reservation changes. This event is triggered by the resource reservation protocol (e.g. RSVP) when a reservation is setup or cancelled.

RESource Link State Advertisements (RES-LSA) advertise the available network resources and are similar to normal Router LSAs[9]. The RES-LSA specify both the available bandwidth and the link delay (milliseconds). Bandwidth is specified by means of a token bucket specification as defined in the Integrated Service architecture [RFC2215] by a bucket depth b (bytes) and a token bucket rate r (bytes/s). A new instance of the RES-LSA is originated whenever there is a new Router LSA or the bandwidth / delay changes significantly. When resource reservation fails due to incorrect information, a router is obliged to advertise its current resources immediately. RES-LSAs are flooded throughout a single area.

Another type of advertisement is the Deterministic Area Border Router Advertisements (DABRA) needed for inter-area multicast QOSPF.

The path selection is based on the network topology, network resource and reservation information and flow QoS requirements (i.e. Tspec of the RSVP Path message). QOSPF calculates the shortest path that satisfies the QoS requirements. Once a QoS path is selected, this path is pinned for the duration of the reservation. This implies that no (better) QoS path is selected unless the existing path can no longer be sustained due to topological changes.

QoS path computation is based on RES-LSAs, Network-LSAs, RRAs and Group-Membership-LSAs. The latter is only used for the multicast case. Links that do not satisfy the QoS requirement are excluded from the path computation. Detailed algorithms to compute the multicast tree for QoS routing are presented [IDOSPF97].

In case of multicast QOSPF both a hop-by-hop routing and source routing (referred to as explicit routing) options is described. Multi-cast paths are computed on demand i.e. triggered by the resource reservation protocol e.g. arrival of RSVP Path message. Path computation is based on a flow i.e. source destination pair (for both unicast and multicast flows). It is noted that the path selected by QOSPF is not necessarily the shortest path.

When there are a large number of flows, the flooding of RRAs may require substantial bandwidth and processing requirements. To make the QoS routing solution scalable, Explicit Routing can be used: only the source router(s) calculate the route and forwarding information is distributed to the downstream routers along the path. This reduces the path computation to a single node. Note that there can be more than one source node when the path travels through different OSPF areas.

4.4.2 Implementation

In [IDOSPF98] an extension to OSPF for support of QoS-based routing is described including detailed algorithms for path computation. Furthermore, experience with an implementation of QOSPF is described. The extensions described are limited to unicast traffic. Both hop-by-hop routing and source routing solutions are discussed.

The goal of this extension is to increase the probability, that a route, that satisfies the QoS requirements, is selected. Furthermore, the aim is to introduce minimum changes to the existing OSPF protocol. It is noted that this implies that optimality is traded off for simplicity.

[9] The RES-LSA information could be encoded in a normal Router LSA by using the special TOS metrics of the Router LSA. However, this would require the increase of the Router LSA advertisements rate.

The QoS requirements are provided by the traffic specification (Tspec) contained in the Path message of the RSVP protocol. Based on this traffic specification and the network state information, the QoS path is selected.

The network state information that is disseminated is limited to bandwidth and delay[10]. This network state information is flooded periodically and when a significant change in the link's available bandwidth occurs. However, QoS path computation only takes into account the bandwidth. Path computation selects the path that satisfies the bandwidth requirement and minimises the number of hops. It is noted that optimising for the bandwidth metric only may result in a delay sensitive traffic flow to be routed along a satellite link with sufficient bandwidth. High latency links could be excluded from path computation based on the applied policy.

Three different stages are identified in QoS-based routing:

1. Transmission of network state information and QoS path selection that meets the QoS requirements of the request.

2. Establishment of QoS paths

3. QoS path management

Path management is concerned with the maintenance of the QoS path. One important issue is that QoS path management should interface with a resource reservation protocol e.g. RSVP. When an RSVP reservation is established, the path is pinned. Finally when the RSVP soft state expires, the path is unpinned.

The frequency with which network state information is updated determines the accuracy of the information upon which QoS path selection is based. There is a trade-off between update frequency and information accuracy. It is recommended to trigger an update whenever a major change in the network state occurs (absolute or relative) with possibly a minimum time between consecutive updates. Furthermore, communication overhead can be reduced when the OSPF flooding mechanism is used unreliably i.e. no acknowledgements are sent. The state information is transmitted so frequently that some updates can be lost without major problems.

Advertising quantised metric values instead of the real metric values, has the advantage that paths with equal QoS arise and these QoS paths may share the traffic load, i.e. load balancing.

There are two basic options when to compute paths: on-demand or pre-computed. The on-demand method computes a path for each request which can be expensive in terms of computational cost. The pre-computed method can divide the computation cost over several requests. However, pre-computation requires path calculation for each possible destination and bandwidth request. Furthermore, depending on the frequency with which pre-computation takes place, inaccuracy is added. Update frequency could be periodic or dependent on the number of updates received. Paths computed in the QOSPF implementation are pre-computed.

The path computation algorithm pre-computes for each destination a minimum hop count path with maximum bandwidth e.g. for each destination the path that provides the most bandwidth in three hops is determined. The computational complexity is comparable to the standard Bellman-Ford shortest path first calculation. It is a property of the Bellman-Ford algorithm that at its h^{th} iteration it identifies the optimal (bandwidth) path between source and destination among paths of h hops. This property is used

[10] The delay parameter is not used yet.

to efficiently compute a routing database from which a minimum hop path with a specified bandwidth can easily be retrieved. This path computation is also referred to as the widest-shortest path.

The QOSPF implementation described above is implemented to evaluate the overhead of QoS-based routing. The overhead involved in QoS path computation, selection of a path from the database, generation of link state update and the processing of link state updates is measured. Furthermore, the processing load for a QoS router in a large network is measured with different pre-computation periods and link state update thresholds. From this analysis it is concluded that the processing cost is within the capabilities of medium-range processor and the transmission cost are tolerable even for large networks as LSA traffic is only a small fraction of the link bandwidth. For further details see [IDOSPF98] and [APOST99]. This implementation has been preceded with a simulation study of the cost associated with different path computation options and update trigger options [APOST98], [APOST98b]. This simulation study included both a high level admission control and different network topologies and traffic models.

4.4.3 Equal Cost Multi-Paths

When there are multiple equal cost paths between source and destination, the OSPF protocol returns a so-called Equal Cost Multi-Path (ECMP), as explained in Section 3.4. Multipaths provide redundancy but could also be used to distribute the load. Several methods of traffic distribution can be applied such as packet round robin, destination address dependent or some hash function on the source-destination address pair[11].

In [IDOSPF99b] it is proposed to use network state information to distribute the traffic over these multipaths. This mechanism can be considered as a limited form of QoS-based routing as QoS requirements are not taken into account, but traffic is routed such that efficient use of network resource is made. This mechanism can also be viewed as dynamic traffic engineering. The network state information is flooded in the network using the Opaque Link State Advertisements (LSA). Opaque LSAs provide a generalised mechanism to allow for extensions of OSPF [RFC2370]. Opaque LSAs consist of a standard LSA header followed by application-specific information. Standard OSPF link-state database flooding mechanisms are used to distribute Opaque LSAs to all or some limited portion of the OSPF topology. The network state information distributed in the network contains: link loading, packet drop rate and link capacity.

4.4.4 Overlay architecture

In [IDOSPF99] an overlay architecture for intra-domain constraint-based routing is presented. Constraint-based routing is defined as a generalisation of QoS-based routing. Constraint-based routing may route flows, flow aggregates or virtual trunks subject to QoS requirements and may take into account either slow or frequently changing network state information. The overlay architecture makes use of the underlying shortest path link state routing protocol. The underlying routing protocol is used to distribute the state information in the network. This approach is different from other proposal where QoS routing functionality is integrated into the routing protocol e.g. QOSPF [IDOSPF97], [IDOSPF98]. The advantage of this approach is that constraint-based routing can be developed independent of the underlying routing protocol using a well-defined interface. Furthermore, the flooding mechanism of link state routing is avoided by implementation of a more efficient method for network state distribution. For this purpose a minimum spanning tree of the network is computed, based on the information provided by the link state routing protocol. This allows a more efficient distribution of network state information compared to flooding where the same network state information may traverse a particular link more than once.

---

[11] A distribution method should preferably preserve the ordering of packets, belonging to a single TCP session, to prevent TCP performance degradation.

4.5 QoS routing and RSVP

To support QoS-based routing, the routing protocol needs to acquire the QoS requirements. A resource reservation protocol, such as RSVP, is a logical choice to provide this information to routing. A routing interface between RSVP and routing is described in Section 4.5.1. An extension to this interface is described in Section 4.5.2. Finally, an interface to RSVP using source routing is evaluated in Section 4.5.3.

4.5.1 Routing interface

The routing interface between RSVP and routing, defined in [IDRSVP98], is used to provide RSVP with the forwarding information to install reservation state within the network. This interface definition supports IPv4 unicast as well as IPv6 unicast and multicast. The routing interface is present in both hosts and intermediate routers.

In [IDRSVP98] it is argued that in case of multicasting, an RSVP capable node does not send out identical copies of an incoming Path messages on all of its outgoing branches (of the multicast tree). One reason could be that the Adspec of the outgoing branches is different (e.g. MTU or latency). Therefore the underlying multicast routing protocol cannot be used in a straightforward manner, but RSVP needs to query the underling routing protocol for outgoing interfaces and generate Path messages for the outgoing interfaces itself. The routing interface includes queries concerning network interface attributes and multi-cast and unicast route specification.

4.5.2 Extension to routing interface

An extension to the interface between RSVP and routing defined in [IDRSVP98] is described in [IDRSVP97]. It is argued that to preserve the independence of RSVP to routing, RSVP could carry opaque routing objects, i.e. carry information relevant to routing. RSVP is able to support different routing schemes such as QoS routing, policy routing, MPLS, etc.

To support QoS-based routing, the reservation protocol needs to provide resource requirements to the routing protocol. The traffic specification (Tspec) of the RSVP path message provides this information. RSVP needs to provide this information to routing when it queries for a route.

Once a QoS path is established, the path management becomes an important issue. The routing protocol should support path management and inform the reservation protocol about path failures. Path pinning prohibits re-routing of the QoS path as load conditions change in the network. However, in some cases unpinning the QoS path is required: RSVP resource reservation fails, or a link failure occurs. Path re-routing can be performed locally at the (failure) node or upstream. When the upstream node needs to re-route the QoS path, this node must be provided the required information (possibly contained in a Path_err message). It is noted that re-routing in case of link failure is more complex and slower with source routing than with hop-by-hop routing. The source node has to be notified of the link failure and re-compute a new route. Hop-by-hop routing can more quickly adapt to the local link failure.

The source routing solution described in [IDRSVP97] is referred to as Explicit Routing. Each RSVP Path message carries an opaque object which specifies the (loose) source route. The information contained in the routing object is made available to routing which returns a modified routing object (pointer adjusted of source route) and the next hop information. There is a similarity with the policy object carried by RSVP where both an outgoing policy object and the policy decision is returned to RSVP.

The following interface extension are proposed in [IDRSVP97]:

Route_Query used by RSVP to find the outgoing interface for sending out the Path message

Route_Reply a reply from routing to a Route_Query.

RSVP_event used by RSVP to notify routing of a specific event concerning a flow (e.g. local reservation failure, receipt of Path_err/Path_tear message, removal of Path state due to soft state timeout, change in PHOP or IP TTL value of Path message (loop detection)).

Routing_event used by routing to inform RSVP of a local link failure.

4.5.3 Source routing

The use of explicit path i.e. source routing and RSVP is discussed in [IDRSVP97b]. The RSVP protocol is extended with an opaque routing element which carries the source routing information. Only unicast routing is considered. The source route may be either strict or loose.

With traditional destination based routing, packets with the same destination follow the same path (the exception to this rule is load balancing). However, with source routing packets destined to the same destination may follow different paths. This allows more control on network resource consumption. Furthermore, with source routing there is a single decision point where with hop-by-hop routing there are multiple decision points which require co-ordination. Therefore loop-free routing is more easily achieved with source routing.

QoS routing includes both the selection of the path that satisfies the QoS requirements of the flow and the guarantee that efficient use of network resources is made. For this purpose QoS routing requires knowledge about the QoS requirements of the flow and network state information. Furthermore, path management (re-routing in case of link failure) and high level admission control policies (rejection of those paths that may satisfy the QoS requirements but are considered too long) are two important issues with QoS routing.

Once a QoS route is established it is preferable that the route remains unchanged i.e. pinned even when routes with better QoS parameters become available. However, when the route can no longer be sustained with the required QoS, re-routing is required. Source routing makes path pinning more easy compared to hop-by-hop routing. Path unpinning is required in (exceptional) cases such as link failures or resource reservation failure. It is noted that when a path is in the process of being pinned, permanent loops can be constructed when the network is in a transient state. The network is in a transient state when state update information is being transmitted and this information has not reached all nodes of the network yet. However, a pinned path with a loop can be detected by considering the TLL field of the Path message. When a Path message with a lower TTL field than the previous Path messages arrives, a loop is detected, and the path should be unpinned.

4.6 Miscellaneous

In this section miscellaneous topics are discussed: Multi-Protocol Label Switching, metric inaccuracy, path selection criteria, multicast routing and multi-class routing. The topics are discussed in more detail in the subsequent sections.

4.6.1 Metric inaccuracy

In general there is inaccuracy involved with the network state information [APOST99b]. There is both a systematic and random inaccuracy. The path computation algorithm can take this inaccuracy into account to improve path computation.

The systematic inaccuracy originates from applying threshold (trigger if there is a large change since last time advertised) or class based (trigger when a class boundary is crossed) triggering policies.

It is possible, given the current value and knowledge of the generation mechanism to compute a probability for the next value. The inaccuracy introduced by the triggering policy can be modeled when the details of the policy are known. Given the current value u and threshold th, the next value n is in the range [u-th, u+th]. This knowledge can be used when paths are computed e.g. apply a so-called 'safety based' path selection (for further details see [APOST99b])

A random inaccuracy originates from delayed updates and clamp down timers. This inaccuracy is difficult to model and predict.

4.6.2 Path selection criteria

In most practical cases there are several feasible paths that satisfy the QoS requirements. When selecting from the set of feasible paths, in general a trade-off is made between network resource consumption and balancing the network load. Selection algorithms put different weight on hop count and load balancing. Several proposed selection criteria are [XIAO99], [MA97b]:

- Widest-shortest path: a path with minimum hop count among feasible paths. If there are several such paths, the one with the maximum reservable bandwidth.
- Shortest-widest path: a path with the maximum reservable bandwidth among feasible paths. If there are several such paths, the one with the minimum hop count.
- Shortest-distance path: a feasible path with the shortest distance. The distance of a k-hop path P is defined as $\mathrm{dist}(P) = \sum_{i=0}^{k} \frac{1}{r_i}$ where r_i is the bandwidth of link i.
- Dynamic-alternative path: Let n be the hop count of a minimum-hop path. A dynamic-alternative path is a widest-shortest path with no more than n+1 hops.

A selection algorithm, other than the shortest-path, consumes more network resources. The widest-shortest path gives high priority to hop count, while shortest-widest path gives high priority to load balancing. The shortest-distance path is a trade-off between the widest-shortest and the shortest-widest path. It favours shortest path when the traffic load is high and widest path when the traffic load is medium to low [XIAO99]. When paths are pre-computed, in principle a different path needs to be (pre-)computed for each bandwidth requirement. Relaxing the single bandwidth requirement to a range of bandwidth requirements, automatically varies between shortest-widest and widest-shortest path selection.

Simulation results indicate that dynamic-alternate path gives better performance when the network is heavily loaded, while balancing the load (e.g. shortest-distance and shortest-widest path) pays off when the load is light [MA97b]. Furthermore, dynamic routing performs significantly better than static routing only when the network load is unevenly distributed, as traffic is re-routed around congested areas.

4.6.3 Multicast Routing

A scalable inter-domain multicast QoS routing solution is presented in [ZAPPA97]. The scalability requirement of inter-domain routing prevents the dynamic distribution of state information in support of QoS-based routing. The solution proposed here is to compute multiple paths for each destination based on relatively static routing metrics. This approach is referred to as Quality of Route (QoR) routing as the metrics indicate static link characteristics such as maximum bandwidth or minimal latency. This routing scheme resembles the OSPF Type of Service routing discussed in Section 3.4 as routes adapt to topological changes and not to resource usage. To increase the probability of establishment of a path that satisfies the QoS requirements, multiple paths are computed. When the primary path fails to satisfy the QoS requirements, an alternate path is selected. Furthermore, the route is pinned after successful establishment.

Many multicast QoS routing schemes employ source routing and require knowledge of the downstream group members. The proposed routing scheme uses these traditional multicast routing schemes to establish pinned QoR multicast paths. These QoR multicast paths are computed decentralised and computation is query driven. The multicast tree is used to build a pinned multicast QoS path. This pinned multicast QoS path is initiated by the receivers i.e. group members. Receivers use (normal) unicast routing to find an explicit path to the sender. A setup protocol installs and pins this route from receiver to the sender and override the traditional multicast tree in the intermediate routers appropriately. For further details see [ZAPPA97].

4.6.4 Multi-class routing

When there are both Best Effort and Guaranteed Service traffic flows in the network, the QoS-based routing of Guaranteed Service flows can have a detrimental effect on the Best Effort traffic (which uses the traditional shortest path algorithm). In a networks with a small amount of Guaranteed Service traffic flows the QoS-path blocking probability is likely to be close to zero irrespective of the QoS-path selection algorithm. However, the effect of the QoS-path selection algorithm on the Best Effort performance can be significant, as is evaluated in [MA98]. This simulation study confirmed that the selection algorithm consuming the most resources has the most detrimental effect on Best Effort routing. This shows that even though the QoS traffic load is light, the effect on Best Effort traffic can be significant.

The routing protocols determine the path between source and destination and therefore to some extent the network resources that are consumed. In [MA99] the network resource sharing between different classes of traffic is evaluated further. In this case the bulk traffic is Best Effort traffic and there is a small portion of QoS traffic. An algorithm, that discourages QoS traffic to select links which are already heavily loaded with Best Effort traffic, is proposed. This is achieved by defining a virtual residual bandwidth per link which indicates the link's residual bandwidth taking into account the congestion conditions of Best Effort traffic. This mechanism adapts automatically to changing traffic mix conditions compared to a scheme where the link's capacity is divided statically for Best Effort and QoS traffic use i.e. static link sharing policies. It is noted that taking into account Best Effort congestion conditions, implies that the strict priority given to QoS traffic, is somewhat relaxed.

5 Traffic engineering in MPLS networks

5.1 Introduction to MPLS

5.1.1 Background

The MPLS technology has emerged from the need to integrate high-speed label-swapping ATM switches into IP routing networks [VISWA98]. MPLS introduces a connection-oriented label switching mechanism in a connectionless IP world. As line speeds increase, the forwarding performance may become a bottleneck in IP routers. ATM switches can operate at much higher speeds than IP routers, therefore better satisfying the endless bandwidth hunger. MPLS is expected to improve the price-performance of network layer routing significantly, and MPLS provides scalability through traffic aggregation.

Many basic MPLS concepts have been borrowed from ATM [DRAKE98]. MPLS and ATM are actually much more alike than they are different. MPLS nodes, which are not based on ATM, will end up looking like ATM switches. Technologies, such as MPLS and DiffServ, are representative of the fact that the best concepts from ATM and traditional IP are merged, to form the Internet of the future.

To run IP over ATM, the IETF has proposed classical IP over ATM, and the ATM forum has proposed LAN Emulation (LANE). Both proposals have the disadvantage that they require $O(n^2)$ connections (with n the number of border nodes). MPLS removes this disadvantage, by integrating IP routing protocols, such as OSPF and BGP, with the label swapping paradigm. Provided that label merging (explained later in the text) is supported, the MPLS proposal scales O(n).

5.1.2 MPLS basics

Multi-Protocol Label Switching (MPLS) proposes a framework for an integrated layer 2 and 3 routing paradigm, referred to as label switching [IDMPLS97], [IDMPLS99]. Packets are routed based on a fixed size label, compared to the traditional IP network layer destination-based routing. The layer 3 protocol can be any of the existing network layer protocols such as IP, IPX, AppleTalk, CLNP etc. Therefore the label switching scheme is referred to as 'multi-protocol'. MPLS is designed to support existing router platforms. MPLS can be implemented in ATM hardware, where the labels are substituted for the VP/VC identifiers. Or in the case of Frame Relay the DataLink Connection Identifier (DLCI) can be used. The intermediate routers in the MPLS domain, referred to as Label Switched Routers (LSR), can be ATM switches or Frame relays.

Within an MPLS domain, packets are routed based on their label, as depicted in Figure 4. The routers at the edge of the MPLS domain are referred to as Label Edge Routers (LER). A LER interworks to the outside world through normal IP routing and labels incoming traffic to the MPLS domain. When there is an incoming IP packet, the LER router examines the destination address of the packet (and possibly other fields) to determine the egress router and to associate a Forwarding Equivalence Class (FEC) with the packet. The FEC defines the way in which packets will be forwarded within the MPLS domain (e.g. path selection and forwarding treatment) [IDMPLS97]. FEC assignment is based on both routing and policy information [IDPOLICY99]. Packets belonging to the same FEC follow the same route and receive the same forwarding treatment within the MPLS domain [IDMPLS97]. The FEC determines the Label Switched Path (LSP) i.e. the path between ingress and egress router in the MPLS domain. Packets with the same FEC, follow the same path. Packets following the same path are also referred to as a traffic trunk. A traffic trunk may contain an aggregated number of flows. Based on the FEC, the packet is assigned a label, which is used inside the MPLS domain to route the packet. When MPLS routing/switching is compared to IP network routing, FEC is done only at the ingress router, while with IP routing FEC is performed at each hop. The label is removed at the egress router.

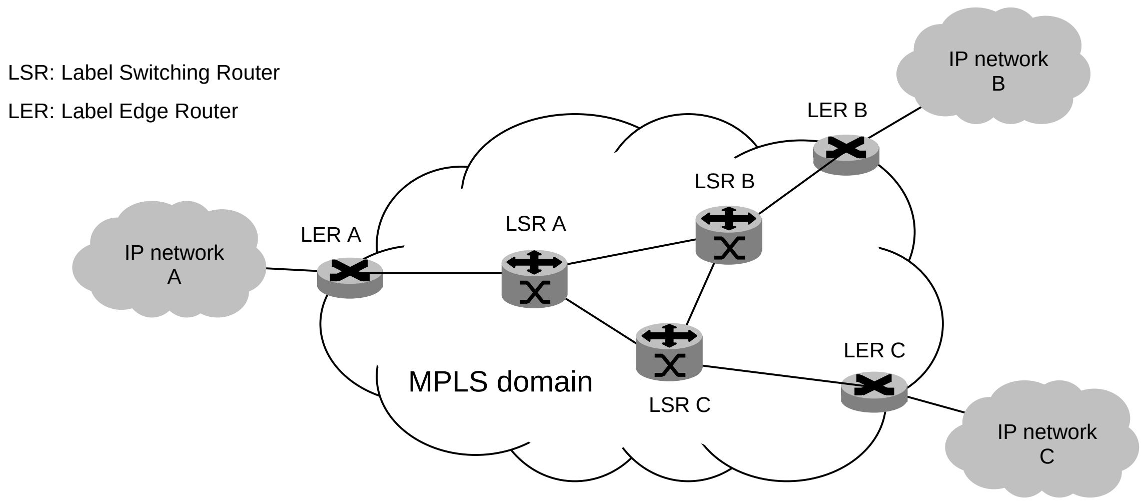

Figure 4 MPLS domain.

In an ATM MPLS domain, an IP packet will undergo segmentation and re-assembly at the ingress and egress node respectively. In IP MPLS domain this is not required.

5.1.3 Labels

The label is either carried in an MPLS-specific 'shim' header or carried in a layer 2 header (e.g. ATM or Frame Relay header). The so-called shim header contains a 20 bits label, 3 bits Class of Service (COS), 1 bit bottom of stack and 8 bits Time To Live. The COS filed allows multiple service classes within the same label. The bottom of stack bit is used for label stacking (explained later in the text). The TTL field, is initialised with the IP TTL value, and decremented with one at each LSR hop.

5.1.4 Label allocation strategies

Label allocation refers to the way in which LSPs are established. LSP establishment includes both label binding (i.e. association of label and FEC) as well as label distribution (i.e. exchanging label bindings with adjacent nodes). Allocation strategies can be divided into traffic-driven (or flow-driven) and control-driven strategies, which will be discussed next.

Traffic-Driven (or Flow-Driven) allocation strategy:

> The traffic-driven allocation strategy is triggered by the actual arrival of user packets. Label allocation is determined by expected traffic patterns (by either matching a well-known port number (to identity the service type associated with the traffic) or monitoring a fixed number of packets to pass by). When a long-lived traffic stream is detected, an LSP is established and the traffic is switched over this LSP. Short-lived traffic patterns may require substantial LSP setup message processing. There is latency involved between traffic arrival and LSP establishment. In this case, first IP routing can be applied, after which label swapping can

be performed, once the LSP is established. The data-driven approach is not limited to a single control domain (e.g. routing domain) and preserves the label space.

With the flow-driven allocation strategy, an LSP is established when a (RSVP) traffic flow is detected. This control strategy requires Multi-Field (MF) packet classification.

Three different types of control-driven label allocations are identified:

1. Topology-Driven

   This allocation strategy is triggered by layer 3 routing control messages, such as routing updates. The label information can be piggybacked in layer 3 control messages. When the topology-driven approach is used, the default LSPs[12] reflect the routing topology and follow the shortest path. These default LSPs are typically used to carry best-effort traffic.

   The label binding and distribution is limited by the size of the network. There is no latency in LSP setup when data arrives, as LSPs are pre-established. The label granularity is equivalent to the advertised routes.

2. Reservation-Driven

   This allocation strategy is triggered by the reception of an RSVP RESV message. The label binding and distribution is limited by the number of flows. Note that the number of flows can be high. The label granularity is equivalent to the traffic flows.

3. Engineering-Driven (or static configuration)

   This allocation strategy is driven by engineering goals. For the purpose of load balancing, different LSPs can be established.

Combinations of triggering policies may exist e.g. topology driven BE and reservation driven QoS traffic.

5.1.5 Label swapping

Traditional IP packet routing is based on the destination address. Therefore a relatively large header needs to be parsed and subsequently a longest match on the network prefix needs to be done (See Section 3.4). Furthermore, some administrative tasks need to be done such as checking the Time To Live (TTL) field and checksum calculation. In case a fixed size label (32 bits) is used, a direct index lookup can be done. The outgoing label and output port can be determined quickly and packet forwarding performance is improved significantly.

Although forwarding is quicker with MPLS, it is not completely free off administrative tasks. For each FEC there is an associated list of procedures, the router will perform on the corresponding packet. This list, denoted as Next Hop Label Forwarding Entry (NHLFE), contains such information as the Next Hop, the operation to perform on the packet's label stack (swap label, pop label, swap and push label(s)) and decrementing the TTL field of shim header. An LSR router examines the label of incoming packets, determines the FEC, executes the associated procedures and queues the packet in the designated output queue. This process is repeated by each intermediate LSR, until the packet reaches the egress LER, where the label information is removed, and the original IP packet is recovered.

[12] Besides these default LSPs, manually configured LSPs or LSPs satisfying certain constraints may be established.

5.1.6 Label binding

In the MPLS architecture, the label binding, i.e. the association between label and FEC, is done by the downstream node[13] [IDMPLS99]. The downstream node informs the upstream neighbouring nodes of the label binding. As labels are only significantly locally, i.e. between two adjacent MPLS nodes, this approach is scalable to large networks.

LSP control refers to the way subsequent label bindings are done along the LSP. Two different types of LSP control are identified: ordered and independent. With ordered LSP control, a node only initiates a label binding if it is the egress node, or if it receives a label binding from a downstream node (a node that is the next hop for that FEC). With independent control, each LSR makes an independent decision to bind a label, and to distribute that binding to its label distribution peers. Independent control may lead to label swapping traffic over a partially established LSP.

An upstream node (e.g. ingress node) may request a binding from its downstream node (downstream-on-demand) or the downstream node may initiate a binding unsolicited (unsolicited downstream).

5.1.7 Label distribution

The distribution of label bindings in the network, can be done with a Label Distribution Protocol (LDP) [IDMPLS99g]. The Label Distribution Protocol allows the construction of Label Switched Paths (LSP) within the MPLS domain. The MPLS architecture is not restricted to a single LDP. Several LDPs have been proposed based on existing (routing) protocols, e.g. IS-IS [IDMPLS99b], OSPF [IDMPLS99c], or new ones [IDMPLS99d].

With LDP two LSRs establish a peer relationship, to exchange label binding information. This LDP adjacency is bi-directional. There are three classes of LDP messages: discovery (announce LSR presence), adjacency (establish, maintain and release adjacencies) and advertisement (advertise label bindings, updates to and withdrawal of bindings) class.

There are two label distribution modes: liberal and conservative. In liberal mode, a node distributes labels bindings to neighbouring nodes, even if these nodes are not part of the LSP. In conservative mode, only those nodes, that are part of the LSP, maintain labels for that LSP. The liberal mode has the advantage that path changes can be effectuated quickly. The conservative mode, on the other hand, conserves a possibly limited label space. The label distribution mode is exchanged during LDP peer initialisation.

The LDP protocol is part of the Component Control of an MPLS node, as depicted in Figure 5. The Component Control (CC) provides both network functionality (e.g. OSPF routing protocol) and label management functions (e.g. LDP).

The Component Control is separated from the forwarding engine, which allows MPLS networks to be build on top of e.g. ATM hardware for switching.

---

[13] A node is denoted as downstream w.r.t. the data stream direction.

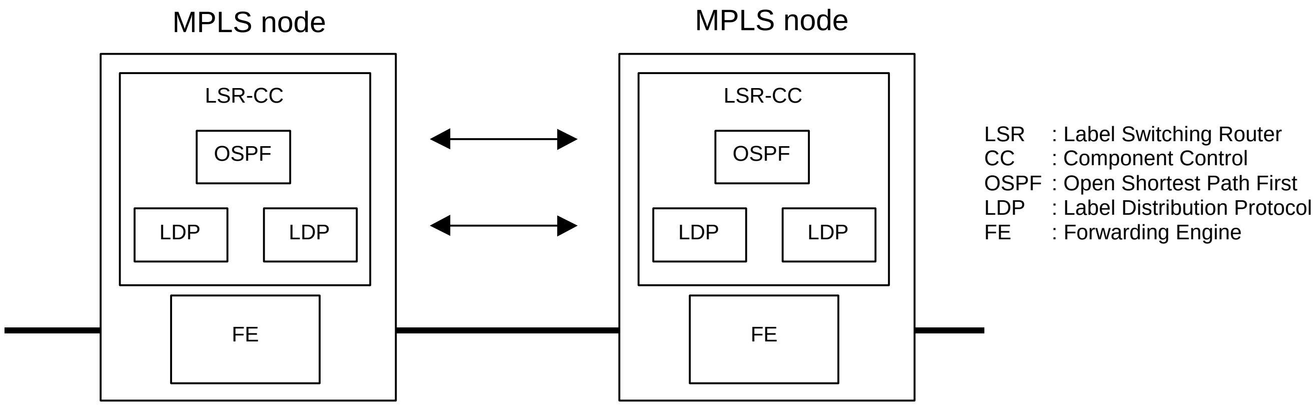

Figure 5 MPLS node architecture.

5.1.8 Example of control-driven LSP setup

The setup of a control-driven LSP is exemplified in Figure 6.

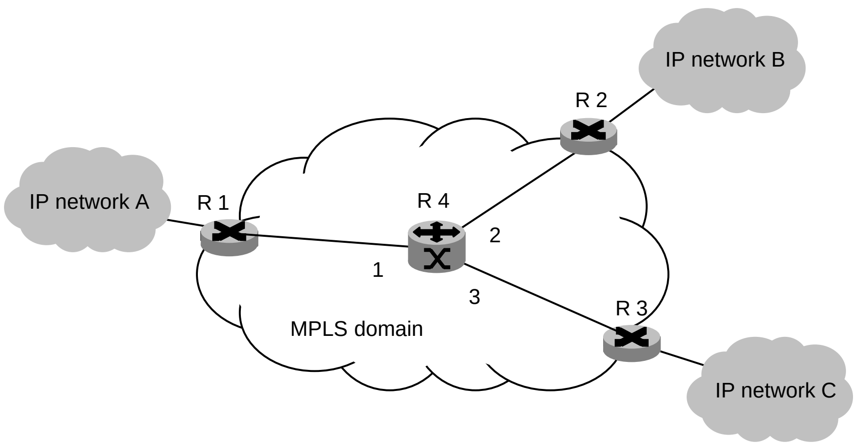

Figure 6 MPLS network configuration example [HAGARD98].

A Forwarding Information Base (FIB), see Table 2, is built by means of IP routing protocols.

Destination	Class of Service	FEC	Next Hop	Outgoing interface
A	X	A1	R1	1
B	X	B1	R2	2
B	Y	B2	R2	2
C	X	C1	R3	3

Table 2 Forwarding Information Base (FIB) of LSR R2.

When the IP routing protocol inserts, deletes or updates an entry in the topological database, an appropriate action is triggered by the Label Distribution Protocol (LDP):

- When a new entry is added, a new LSP is established.
- When an entry is deleted, the corresponding LSP is released.
- When an updated entry indicates a new route to the destination, the LSP is re-established.

By means of the LDP protocol, the Label Information Database (LIB) is build, see Table 3.

FEC	Incoming Label	Incoming interface	Outgoing Label	Outgoing interface
A1	BL1	2	AL1	1
A1	CL1	3	AL2	1
B1	AL1	1	BL1	2
B2	AL2	1	BL2	2
B1	CL2	3	BL3	2
B2	CL3	3	BL4	2
C1	BL2	1	CL1	3
C1	AL3	2	CL2	3

Table 3 Label Information Base (LIB) of LSR R2.

During the life-time of the LSP, the LDP protocol is responsible for the maintenance of the LIB.

5.1.9 Loops

Loops arise when an LSP is established which contains a loop. ATM cells do not contain a TTL field to limit the damage done by loops. However standard ATM PNNI routing ensures that no routing loops are established. When routing is based on IP routing information, transient loops may arise, which will be eliminated after some time when IP routing converges. However, congestion caused by routing loops, may increase the routing convergence time.

Three different loop handling methods are identified:

- Loop survival minimises the effect of loops. For this purpose the TTL value is decremented with one at each hop and, packets with TTL zero are discarded. Short term transient loops normally are repaired by the routing protocol when it converges. It is noted that fair queueing can prevent 'looping' traffic to negatively influence other traffic flows, when separate queues are employed for these flows.
- Loop detection eliminates loops once they are detected. In general, the TTL field provides the ability for loop detection.
- Loop prevention prohibits (layer 2) loops to be established.

5.1.10 Hierarchical routing and label stacking

Label stacking allows MPLS to provide end-to-end services across different domains, using hierarchical routing, as depicted in Figure 7. A packet may contain multiple labels when routed across different domains.

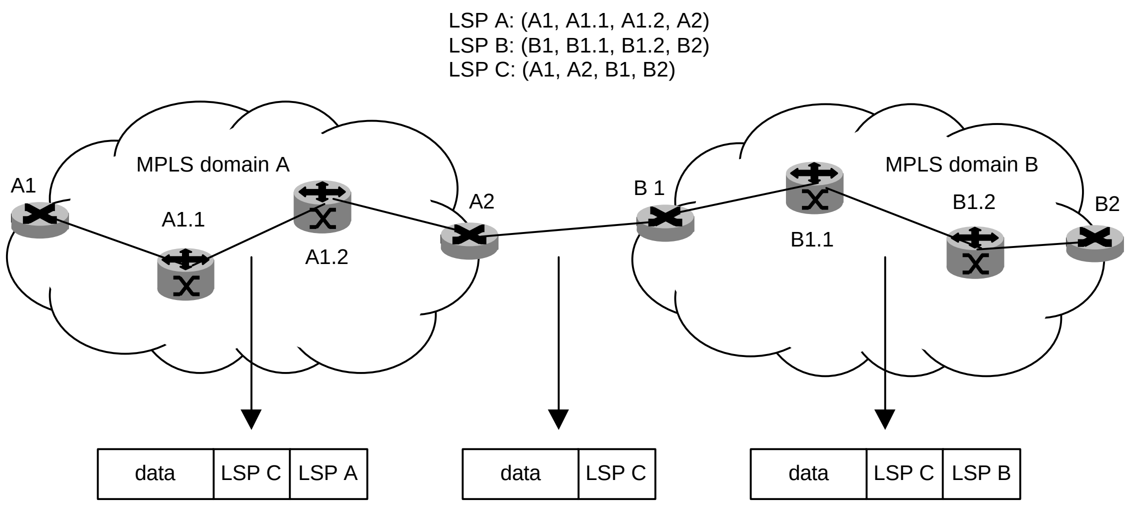

Figure 7 Hierarchical routing and label stacking.

The labels are pushed and popped at the ingress and egress border routers respectively. Label stacking provides similar functionality as IP-in-IP tunneling and loose source routing.

Label stacking can also be used to maintain the identity of different streams, when they are aggregated into a single LSP. The MPLS multi-level hierarchy is similar to the dual level hierarchy of VP tunnels in ATM.

5.1.11 Granularity

A flow refers to the application traffic between hosts. A flow is defined end-to-end between hosts. A trunk refers to the aggregated traffic of flows of the same class. The aggregated traffic flows follow the same route within the MPLS domain. Trunks are routable objects like virtual circuits in ATM. A Label Switched Path (LSP) refers to the aggregated traffic, that belongs to the same forwarding equivalence class. An LSP can contains multiple traffic trunks of different classes. Both a trunk and an LSP are defined within an MPLS domain. A link is the medium that carries the information. A link is defined locally.

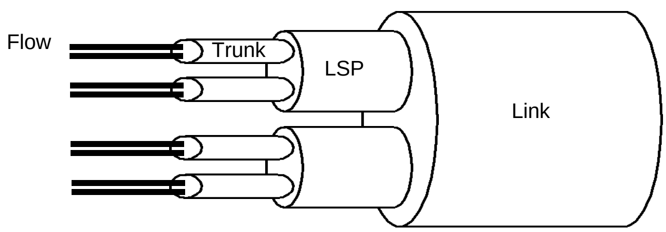

Figure 8 Relationship link, LSP, trunk and flow [BHANI99].

MPLS allows different forwarding granularities, ranging from coarse (e.g. traffic destined to a particular egress router) to fine grain (e.g. application traffic flows between hosts).

5.1.12 LSP types

Four types of label switched paths are identified:

- point-to-point (in small networks between each ingress and egress router (scales $O(n^2)$!), or for application flows between hosts e.g. RSVP flows)
- multipoint-to-point (in large networks to scale O(n), requires merging capabilities of the interior MPLS nodes, which is explained in the following text)
- point-to-multipoint (to support multicast traffic)
- multipoint-to-multipoint (multicast traffic with different sources)

MPLS scalability is provided by two features: forwarding based on an inverted tree (spanning tree rooted at the destination node) and reduction of the number of destinations by means of aggregation. When an MPLS node is capable of merging, O(n) switched multipoint-to-point paths provide reachability to n destinations.

5.1.13 Merging

Label merging is the process of labelling packets, with the same FEC but different incoming labels, with the same outgoing label. Label merging is required to support multipoint-to-point LSPs.

In ATM, data packets are encapsulated into an ATM Adaptation Layer, say AAL5, and the AAL5 PDU is segmented into ATM cells. Each ATM cell is identified by a VPI/VCI value and ATM cells are transmitted in sequence. ATM switches are required to keep the cells of a PDU contiguous and in sequence. This is because the device, that re-assembles the cells into PDUs expects the cells to be contiguous and in sequence. There is not sufficient information contained in the ATM cell header (unlike IP fragmentation) to reassemble the PDU with any cell order. Hence, if cells from several upstream links are transmitted onto the same downstream VPI/VCI, then cells from one PDU can get interleaved with cells from other PDUs. This results in corruption of the original PDUs at re-assembly.

There are two options to support merging in MPLS over ATM: VC merge and VP merge:

When VC merge is applied, different incoming VCs are merged into a single outgoing VC. This requires the LSR either to perform SAR (i.e. re-assembly of AAL5 frames) or to buffer cells. When SAR is applied, complete ATM frames are re-assembled before they are merged on the outgoing VC. Complete re-assembly can be avoided, when the cells are buffered and the begin- and end-of-frame cells can be identified in the buffer. It is noted that VC merge can deteriorate the effect of traffic shaping.

With VP merge, different incoming VPs are merged into a single outgoing VP. Frames from different sources, within a single VP, are identified by their VCI. It should be noted that VP merge preserves traffic shaping.

VC merging has not yet been standardised. With current ATM technology point-to-point LSPs are used. The introduction of VC merging will increase the scalability significantly but will require a hardware update.

5.1.14 Unlabeled traffic

In general the MPLS carries labelled traffic, however, in some situations unlabelled traffic may enter [HAGARD98]:

- When the LER has not yet established an LSP, but already forwards traffic.
- When an outgoing MPLS interface fails, outgoing packets are transmitted unlabelled.
- When aggregated traffic leaves the network through different egress routers, layer 3 routing is required when the aggregated traffic is de-aggregated.

All unlabelled traffic runs over a default VPI/VCI ATM connections. For incoming unlabelled packets a layer 3 routing table lookup is performed to determine the outgoing interface.

5.1.15 Overview

Now that the basics of MPLS have been discussed, the QoS routing capabilities provided by MPLS, can be evaluated. In MPLS QoS-based routing, is referred to as constraint-based routing. MPLS constraint-based routing is discussed in Section 5.3. The construction of explicit routes, discussed in the Section 5.4, is an important feature of constraint-based routing. Finally the use of RSVP in MPLS is discussed in Section 5.5. Constraint-based routing is part of a larger notion, referred to as Traffic Engineering in MPLS. MPLS Traffic Engineering is therefore discussed first in Section 5.2.

5.2 Traffic Engineering

The goal of Traffic Engineering (TE) is to facilitate efficient and reliable network operations, and optimise the utilisation of network resources [IDMPLS98]. TE objectives can be divided into traffic oriented and resource oriented objectives. Traffic oriented objectives aim to improve the QoS characteristics of traffic streams (e.g. packet loss, delay, delay variation and throughput). Resource objectives refer to the efficient use of network resources, especially bandwidth. Resource objectives should prevent congestion in one part of the network, while other parts of the network, providing alternate paths, are under-utilised.

Minimising congestion is both a traffic and a resource objective. Congestion, in this context, does not refer to short term transient congestion due to traffic bursts, but refers to long term periods, where QoS requirements are not met. Congestion can arise from insufficient resources or inefficient mapping of traffic to network resources. The former cause must be addressed with the installation of additional resources or with the use of congestion control techniques (e.g. rate limiting, flow control, queue management, etc). The latter cause can be addressed with Traffic Engineering, as there are sufficient resources. One important technique employed by Traffic Engineering is load balancing. Load balancing aims to minimise maximum resource utilisation.

It should be noted that, in general, Interior Routing Protocols (IRP) do not provide efficient Traffic Engineering control functions. IRP routing protocols frequently route traffic over the shortest path, without considering the bandwidth requirements. Thus traffic is routed over paths which may not provide sufficient resources. Some limited control is provided by the use Equal Cost Multi-Paths (ECMP), see Section 4.4.3, however routing traffic over links with insufficient resources is not prohibited.

It should be noted that in case IP over ATM is employed, the underlying layer 2 provides several TE control functions: constraint based routing (PNNI), provisioning of VC paths, call admission control, traffic shaping and policing.

A traffic trunk is defined as an aggregation of traffic flows of the same class. A traffic trunk is mapped on an LSP for transmission. Traffic Engineering can be viewed as mapping the MPLS graph, consisting of LSRs connected by LSPs between them, onto the topology graph representing the real network topology [IDMPLS98]. The edges in both the MPLS graph and the topology graph have attributes, expressing the trunk characteristics and the link characteristics respectively.

The following trunk attributes are identified [IDMPLS98]: traffic parameter, generic path selection and maintenance, priority, pre-emption, resilience and policing attributes.

> The *traffic parameter* attribute captures the traffic characteristics of the traffic stream.
>
> The *generic path selection and maintenance* attributes provide means to enable policy control such as: enable administratively specified explicit paths, defining preference rules for selection of explicit paths, defining so-called resource class affinity (which specifies which resource classes to include in or exclude from constraint based path selection), adaptability of path to changing network conditions (similar to pinning or unpinning paths) and load distribution across parallel traffic trunks. For further detail see [IDMPLS98].
>
> The *priority* attributes defines the relative importance of traffic trunks. This attribute allows high priority traffic to be routed along pre-determined paths even under high load conditions. High priority traffic causes re-routing of lower priority traffic.
>
> The *pre-emption* attribute of a traffic trunk specifies whether this traffic trunk can pre-empt (i.e. re-route) another traffic trunk, and whether this traffic trunk can be pre-empted by another traffic trunk.
>
> The *resilience* attribute defines the behavior of the trunk under fault conditions.
>
> The *policing* attribute defines the actions to be taken when the traffic of the trunk does no longer comply to the traffic characteristics specified by the traffic attribute.

Two link attributes are identified: maximum allocation multiplier and resource class attribute. The maximum allocation multiplier specifies the proportion of the available link resources that can be assigned to traffic trunks. The resource class attribute allows resources to be divided into different classes. Different policies concerning resource utilisation can be applied to different classes.

The mapping of the MPLS graph to the topology graph is obtained through constraint-based routing. The constraint based routing protocol takes as input the MPLS graph, topology graph, link attributes and resource attributes (and possibly other topology information). Based on this information, an (explicit) route is determined for the traffic trunk, that satisfies the trunk attributes. A simple constraint based routing algorithm, as an example, is described in [IDMPLS98]. In this example, links are pruned from the topology graph, which do not satisfy the requirements of the traffic trunk attributes. Next, the resulting graph is used to determine the shortest path. A more elaborate example of constraint-based routing is discussed in the following section.

5.3 Constraint-Based routing

Extensions to the LDP protocol, to support constraint-based routing, are presented in [IDMPLS99f]. Detailed objects and procedures are defined, that enable the setup of constraint-based LSPs. The constraint-based routing proposal discussed in here, is actually a refinement of the constraint-based routing concept, discussed in the previous section.

Constraint-based routing, in this context, includes QoS-based routing, but is a broader notion. Besides QoS requirements other constraints, such as policy constraints, may apply.

Both strict and loose explicit routes can be specified (see Section 5.4). Explicit routes are specified as a series of hops.

Traffic parameters are used to specify the characteristics of the traffic, carried along the constraint-based route. The maximum allowed peak rate of the traffic flow is specified by a token bucket specification, denoted by Peak Data Rate (PDR) (bytes/sec) and Peak Burst Size (PBR) (bytes). The committed data rate specifies the amount of resources, that are committed to be available for this traffic stream in the MPLS domain. The committed data rate is specified by a token bucket specification, denoted by Committed Data Rate (CDR) (bytes/sec) and Committed Burst Size (CBS) (bytes).

The constraint-based routing proposal includes the definition of a pre-emption, route pinning and resource class object, as specified in the previous section. The constraint-based routing proposal makes use of existing routing protocols, to disseminate network state information. Furthermore explicit routes are used (see Section 5.4), but details concerning the routing algorithm are not included in the document.

5.4 Explicit routing

Route selection refers to the method of selecting an LSP for a particular FEC. There are two options: hop-by-hop or explicit routing. Hop-by-hop routing allows each node to select the next hop for each FEC, similar to IP hop-by-hop routing. With explicit routing, there is a single node (possibly egress or ingress node), that determines the LSP. An explicit route is defined as a route, which is explicitly specified as a sequence of hops. Explicit routing is more efficient than IP source routing in terms of packet overhead. Once the explicit path is established, each packet only carries the label. With IP source routing, each IP packet carries the source routing information. Both strict and loose source routing are supported.

To setup either a point-to-multipoint as well as a multipoint-to-point LSP, a (minimum spanning) tree is required. In [IDMPLS99e] an explicit tree routing object is defined, which contains the specification of such a tree. This routing object, referred to as explicit tree routing object, can be inserted into LDP messages to establish an LSP. The explicit tree route can be obtained from dynamic routing protocols or manually configured. The explicit tree route is specified with a common <type, length, value> specification method.

Explicit control over the routing of LSPs can be required for both policy and network efficiency reasons (e.g. load balancing). Explicit routing provides this control. An explicit route is initiated by the LER and establishes an Explicitly Routed LSP (ER-LSP). Both LDP and RSVP support explicit routing. The RSVP support for explicit routing is discussed in more detail in the following section.

5.5 Using RSVP

In this section the use of RSVP, within an MPLS domain to setup explicit routes, is evaluated. To setup explicit routes with RSVP, an Explicit Route Object (ERO) is defined [IDMPLS97b]. The ERO object contains the route specification i.e. sequence of hops. Furthermore, LABEL objects may be inserted into RSVP messages, which carry the label binding information. This allows the establishment of Label Switched Path (LSP) along the explicit route. This type of LSP is referred to as Explicitly Routed LSP (ER-LSP). The RSVP protocol could also be used to reserve resources along the LSP, but resource reservation is not required. An ER-LSP could, for example, be used to carry Best Effort traffic.

The establishment of an ER-LSP is exemplified in Figure 9. The Label Edge Router (LER) A creates an RSVP Path message, and inserts an Explicit Route Object (ERO) and LABEL_REQUEST object. The Path message is then routed along the path, specified by the ERO object (thick lines). Each intermediate router, along the path, installs Path state (e.g. reverse routing state). Upon receipt of the Path message by the destination node LER C, the LABEL_REQUEST object is detected, and LER C initiates the setup of an LSP along the explicit route, specified by the ERO object. LER C creates an RSVP Resv message and inserts a LABEL object, specifying the label binding. The Resv message is send upstream to LER A, using the installed reverse routing state. Each intermediate node inspects the LABEL object, and updates its local label binding for the node upstream. As a result an LSP is established along the explicit route. Resources may have been reserved for the LSP, along the normal RSVP procedures, but this is not explicitly described in [IDMPLS97b].

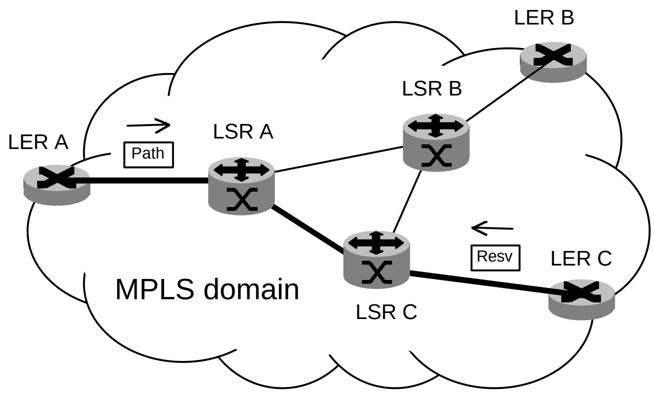

Figure 9 Explicit Route LSP example.

The decision to route traffic along an ER-LSP is taken by the Label Edge Router (LER). This decision can be based on the RSVP filterspec, but other rules may apply. Only the ingress router is concerned with the mapping of incoming traffic to explicit routes. In case RSVP flows are mapped onto ER-LSP, this implies that a reservation state can be identified by a fixed size label, instead of the standard Multi-Field (MF) classification used with RSVP [IDMPLS98d].

MPLS can be used to create aggregate tunnels for RSVP flows. In this way the amount of state, to be stored in the core network, is reduced, and the RSVP scalability is improved. However in [IDMPLS98c] it is proposed to use a different MPLS label for each RSVP flow. This implies that for each flow an LSP is established. It is noted that a different label for each sender is required, when there are multiple senders within the RSVP session (unless a Wildcard Filter (WF) reservation style is used where de-aggregation of sender flows at branch points is not required).

Different RSVP objects have been defined for the purpose of ER-LSP establishment and maintenance [IDMPLS98d]:

When during an RSVP session, the sender finds a better route, the explicitly established route can be re-routed on the fly by changing the EXPLICIT_ROUTE object of the Path message.

The RECORD_ROUTE object allows the sender to obtain the exact path, followed by the Path messages. This allows the sender to monitor any changes in the path.

A SESSION_ATTRIBUTE object is added for session identification and diagnostics including pre-emption, priority and fast re-route.

The EXPLICIT_ROUTE objects allows the definition of an explicit path along which the Path messages should travel.

The LABEL_REQUEST object request label bindings to be established along the path followed by the Path messages.

6 Conclusions

6.1 Final thoughts

QoS-based routing can improve the probability of successful establishment of a path, satisfying the QoS requirements. However, this improvement comes at a cost. Whether the benefits outweigh the costs is a multi-facet problem. There is an obvious reluctance to introduce QoS-based routing solutions in the Internet, as it introduces additional complexity.

The deployment of QoS-based routing will increase the dynamics of path selection. This will make the path selection of network traffic less predictable. It is expected that network management becomes more complex with QoS-based routing. Sufficient policy control functions should be provided by QoS-based routing solutions, to provide the necessary control for network management.

The Integrated Services architecture, with its flow-based reservations, is not scalable to the core of the Internet. Therefore QoS-based routing solutions, which establish QoS paths for flows, such as QOSPF, are also not applicable for the core of the Internet. The Differentiated Services concept, with its scalability and simplicity, can provide QoS in the core of the Internet. Furthermore, the traffic engineering features of MPLS can also provide QoS-based routing of trunk traffic in the core of the Internet.

In case Integrated Services are deployed in the access part of the network, flow-based QoS routing can further enhance the service offering. However, when QoS-based routing is deployed without a resource reservation mechanism, it becomes more complicated to sustain the QoS requirements, after the QoS path has been established.

6.2 Trends

An attempt is made to point out trends in the development of QoS-based routing and identify potential future directions. It should be noted that these are solely the reflections of the author.

Much effort concerning QoS-based routing has been performed in the Quality of Service routing working group of the IETF. However, this working group has been closed. One reason why the working group is closed is that the working group has achieved its goal: to define a framework and techniques for Quality of Service (QoS) Routing in the Internet i.e. RFC 2386. Another reason is that an overall QoS-based routing solution is considered impractical, including both inter- and intra-domain routing. Research has shifted towards QoS-based routing for specific routing domains e.g. QOSPF and MPLS.

The QOSPF proposals are discussed in the OSPF working group obviously. The QOSPF proposal of Apostopoulos et all [IDOSPF98] is detailed and the applicability has been shown through an implementation. Given that the Internet community views this as a viable option, further research and implementation trials in this direction can be expected. However, the QOSPF proposals are tied to the Integrated Service concept which lacks scalability to large networks.

The MPLS proposals seem to gain momentum as major router vendor starting to offer MPLS features. Further research in this area can be expected.

As the Internet has moved from a research network to a network with significant commercial potential, it can be expected that work within the IETF is vulnerable to business interests and political issues.

Technical issues may become blurred with commercial interests. No doubt QoS-based routing proposals will be subject to this scrutiny.

6.3 Internet Next Generation project

In this section QoS-based routing is evaluated in the light of the Internet Next Generation project.

The Internet Next Generation (ING) project is a research project in co-operation with major Dutch R&D facilities and the University of Twente. The project aims to contribute to the development of IP-based solutions to provide enhanced Quality of Service. Furthermore, mechanisms to support user mobility and wireless applications are considered. Research is focussed on solutions at the network layer. For further detail see [ING99].

Research, carried out by Ericsson, in the ING project concentrates on the (wireless) access network. Deployment of a QoS-based routing solution in the access network provides an improvement given that there is congestion. However, the access network is likely to provide sufficient resources, reducing the benefit of QoS-based routing. Furthermore, the selection of QoS routing solution in the access network depends on the QoS architecture that is deployed in the access network: Integrated Service, Differentiated Service or MPLS. When the architecture choice is clear, the need for a QoS-based routing solution can be discussed further.

6.4 For further study

When QoS-based routing is applied in wireless access networks, two new issues arise. First of all the quality of the airlink may vary, in which case the QoS-path must be re-routed when the QoS commitments can no longer be sustained. Possibly QoS information is passed from the wireless datalink layer to the higher routing layer, to indicate a need for re-routing. Secondly the mobility of the users makes re-routing necessary when the user moves to a new point of attachment. The required resources to re-route a QoS path, may possibly not be available at the new point of attachment. Therefore special attention should be paid to the selection of the new point of attachment, when there are multiple possible points of attachment. To provide seamless handovers, multiple paths may be established to provide un-interrupted data transfer. Furthermore, stringent requirements may be imposed on the delay involved with handover.

An unresolved issue concerning QoS-based routing concerns the timescale over which network state changes in the Internet. A QoS-based routing scheme is concerned with the distribution of network state information. In case the network state changes thus frequently that either the distribution overhead becomes unacceptable or the disseminated information is out of date, then a QoS-based routing solution becomes impractical. Studies concerning Internet traffic characteristics have revealed self-similar behavior [PAXSON95]. One significant aspect of self-similar traffic is that the variance is independent of the time-scale. Further study of the validity of this concern is needed[14].

The support of multicast QoS trees with receiver heterogeneity and shared reservation styles is still an unresolved issue in many QoS-based routing solutions.

When no specific policy constraints are implemented and admission control is solely based on the available resources, flows requesting a small amount of resources are more likely to succeed compared to flows requesting a large amount of bandwidth. This introduces the issue of fairness. A definition of fairness should be provided and policy constraints that effectuate this fairness should be evaluated.

---

[14] When the QoS metric is the unreserved bandwidth, then this metrics is assumed to vary slowly compared to the link utilisation (which include traffic variation of all traffic).

In many cases the QoS path computation is based on the sender traffic characteristics made available to QoS routing by the RSVP protocol's Tspec. The required bandwidth is computed based on the sender Tspec. However, the actual required resources are computed by the receiver e.g. bandwidth specified in Rspec. The amount of resources calculated by the QoS-based routing protocol and reservation protocol are not necessarily the same resulting in a failed QoS connection setup.

7 References

7.1 By alphabet

[AF96c] Title: Private Network-Network Interface Specification
Author: ATM forum
Doc. no.: Version 1.0
Date: March 1996

[APOST98] Title: Quality of Service Based Routing: A Performance Perspective
Author: G. Apostolopoulos, R. Guerin, S.K. Tripathi
Doc. no Computer Communication Review, Volume 28, Number 4
Date: October 1998

[APOST98b] Title: On Reducing the Processing Cost of On-Demand QoS Path Computation
Author: G. Apostolopoulos, R. Guerin, S. Kamat, S.K. Tripathi
Doc. no Internation Conference on Networking Protocols (ICNP)
Date: October 1998, Austin, Texas

[APOST99] Title: Implementation and Performance Measurements of QoS Routing Extensions to OSPF
Author: G. Apostolopoulos, S. Kamat, R. Guerin
Doc. no INFOCOM'99, New York
Date: March 21-25, 1999

[APOST99b] Title: Improving QoS Routing Performance Under Inaccurate Link State Information
Author: G. Apostolopoulos, R. Guerin, S. Kamat, S.K. Tripathi
Doc. no Proceedings of the 16th International Teletraffic Conference (ITC)
Date: June 1999, Edingburgh

[ASH98] Title: Dynamic Routing in Telecommunications Networks
Author: G.R. Ash
ISBN: 0070064148 (McGraw-Hill)
Date: 1998

[BHANI99] Title: Quality of Service using Traffic Engineering over MPLS: An Analysis
Author: P. Bhaniramka, W. Sun, R. Jain
Doc. no.: Globecom'99
Date: March 1999

[CIDON97] Title: Multi-Path Routing combined with Resource Reservation
Author: I. Cidon, R. Rom, Y. Shavitt
Doc. no.: Infocom'97
Date: April 1997, Kobe, Japan

[DRAKE98] Title: MPLS/ATM integration (presentation)
Author: J. Drake
Doc. no.: http://info.uu.net/ads/techconf/presentations/
Date: November 1998, UUNET IW-MPLS'98 Conference

[PATENT99] Title: Ericsson Patent Search
Author: Derwent
Doc. no.: EPS V3.1
Date: June 1999

[GUERIN97] Title: QoS-based Routing in Networks with Inaccurate Information: Theory and Algorithms
Author: R. Guérin, A. Orda
Doc. no.: Infocom'97
Date: April 1997, Kobe, Japan

[HAGARD98] Title: Multiprotocol label switching in ATM networks
Author: G. Hagard, M. Wolf
Doc. no.: Ericsson Review, No. 1
Date: 1998

[IDMPLS97] Title: A Framework for Multiprotocol Label Switching
Author: R. Callon, P. Doolan, N. Feldman, A. Fredette, G. Swallow, A. Viswanathan
Doc. no.: Internet Draft (work in progress)
Date: 21 November 1997

[IDMPLS97b] Title: Explicit Route Support in MPLS
Author: B. Davie, T. Li, E. Rosen, Y. Rekhter
Doc. no.: Internet Draft (work in progress)
Date: November 1997

[IDMPLS98] Title: Requirements for Traffic Engineering Over MPLS
Author: D.O. Awduche, J. Malcolm, J. Agogbua, M. O'Dell, J. McManus
Doc. no.: Internet Draft (work in progress)
Date: October 1998

[IDMPLS98b] Title: Framework for Traffic Management in MPLS Networks
Author: P. Vaananen, R. Ravikanth
Doc. no.: Internet Draft (work in progress)
Date: March 1998

[IDMPLS98c] Title: Use of Label Switching With RSVP
Author: B. Davie, Y. Rekhter, E. Rosen, A. Viswanathan, S. Blake
Doc. no.: Internet Draft (work in progress)
Date: March 1998

[IDMPLS98d] Title: Extensions to RSVP for Traffic Engineering
Author: D. Awduche, D. Gan, T. Li, G. Swallow, V. Srinivasan
Doc. no.: Internet Draft (work in progress)
Date: August 1998

[IDMPLS99] Title: Multiprotocol Label Switching Architecture
Author: E.C. Rosen, A. Viswanathan, R. Callon
Doc. no.: Internet Draft (work in progress)
Date: April 1999

[IDMPLS99b]	Title:	IS-IS extensions for Traffic Engineering
	Author:	H. Smit, T. Li
	Doc. no.:	Internet Draft (work in progress)
	Date:	February 1999
[IDMPLS99c]	Title:	OSPF Extensions for Traffic Engineering
	Author:	D.M. Yeung
	Doc. no.:	Internet Draft (work in progress)
	Date:	February 1999
[IDMPLS99d]	Title:	LDP specification
	Author:	L. Andersson, P. Doolan, N. Feldman, A. Fredette, B. Thomas
	Doc. no.:	Internet Draft (work in progress)
	Date:	May 1999
[IDMPLS99e]	Title:	Explicit Tree Routing
	Author:	H. Hummel, S. Loke
	Doc. no.:	Internet Draft (work in progress)
	Date:	February 1999
[IDMPLS99f]	Title:	Constraint-Based LSP Setup using LDP
	Author:	B. Jamoussi (editor)
	Doc. no.:	Internet Draft (work in progress)
	Date:	February 1999
[IDMPLS99g]	Title:	LDP specification
	Author:	L. Andersson, P. Doolan, N. Feldman, A. Fredette, B. Thomas
	Doc. no.:	Internet Draft (work in progress)
	Date:	June 1999
[IDOSPF97]	Title:	Quality of Service Extensions to OSPF or Quality Of Service Path First Routing (QOSPF)
	Author:	Zhang, Sanchez, Slakewicz, Crawley
	Doc. no.:	Internet Draft (work in progress)
	Date:	September 1997
[IDOSPF98]	Title:	QoS Routing Mechanisms and OSPF Extensions
	Author:	G. Apostolopulos, R. Guerin, S. Kamat, A. Orda, T. Przygienda, D. Williams
	Doc. no.:	Internet Draft (work in progress)
	Date:	14 April 1998
[IDOSPF99]	Title:	An Overlay Model for Constraint-Based Routing
	Author:	B. Rajagopalan, Q, Ma
	Doc. no.:	Internet Draft (work in progress)
	Date:	20 Januari 1999
[IDOSPF99b]	Title:	OSPF Optimized Multipath (OSPF-OMP)
	Author:	C. Villamizar
	Doc. no.:	Internet Draft (work in progress)
	Date:	24 Januari 1999

[IDPOLICY99] Title: Terminology for describing network policy and services
Author: J. Strassner, E. Ellesson
Doc. no.: Internet Draft (work in progress)
Date: 23 June 1999

[IDROUTE98] Title: Routing of Multimedia Connections when Internworking with PSTN, ATM and IP Networks
Author: G.R. Ash, Y. Lee
Doc. no.: Internet Draft (work in progress)
Date: November 1998

[IDRSVP97] Title: Extended RSVP-Routing Interface
Author: R. Guerin, S. Kamat, E. Rosen
Doc. no.: Internet Draft (work in progress)
Date: 29 July 1997

[IDRSVP97b] Title: Setting up Reservations on Explicit Paths using RSVP
Author: D.H. Gan, R. Guerin, S. Kamat
Doc. no.: Internet Draft (work in progress)
Date: 21 November 1997

[IDRSVP98] Title: RSRR: A Routing Interface For RSVP
Author: D. Zappala
Doc. no.: Internet Draft (work in progress)
Date: 1 July 1998

[ING99] Title: Project Proposal Telematics Institute; Internet Next Generation
Author: A. Prass (editor)
Doc. no.: http://ing.ctit.utwente.nl/background/public.pdf
Date: 27-01-1999

[GAREY79] Title: Computers and Intractability: A Guide to the Theory of NP-completeness
Author: M.R. Garey, D.S. Johnson
Publisher: W.H. Freeman and Company
Date: 1979

[MA97] Title: Quality-of-Service Routing for Traffic with Performance Guarantees
Author: Q. Ma, P. Steenkiste
Doc.no.: IFIP 5th Int. Workshop on Quality of Service, Columbia University, New York
Date: 1997

[MA97b] Title: On Path Selection for Traffic with Bandwidth Guarantees
Author: Q. Ma, P. Steenkiste
Doc.no.: International Conference on Network Protocols, Atlanta
Date: October 1997

[MA98] Title: Routing Traffic with Quality-of-Service Guarantees in Integrated Service Networks
Author: Q. Ma, P. Steenkiste
Doc.no.: Workshop on Networking and Operating Systems Support for digital audio and video

	Date:	July 1998
[MA99]	Title:	Supporting Dynamic Inter-Class Resource Sharing: A Multi-Class QoS Routing Algorithm
	Author:	Q. Ma, P. Steenkiste
	Doc.no.:	IEEE Infocom'99
	Date:	March 1999
[MOY98]	Title:	OSPF; Anatomy of an Internet Routing Protocol
	Author:	J.T. Moy
	Doc.no.:	0-201-63472-4
	Date:	1998
[ORDA99]	Title:	Routing with End to End QoS Guarantees in Broadband Networks
	Author:	A. Orda
	Doc.no.:	Transactions on Networking
	Date:	June 1999
[PAXSON95]	Title:	Wide Area Traffic: The Failure of Poisson Modeling
	Author:	V. Paxson, S. Floyd
	Doc. no.:	IEEE/ACM Transactions on Networking
	Date:	June 1995
[PAXSON97]	Title:	End-to-end Routing Behavior in the Internet
	Author:	V. Paxson
	Doc. no.:	IEEE/ACM Transactions on Networking, Vol.5, No.5, pp. 601-615
	Date:	October 1997
[RFC1584]	Title:	Multicast extensions to OSPF
	Author:	J. Moy
	Doc. no.:	RFC 1584
	Date:	March 1994
[RFC1633]	Title:	Integrated Services in the Internet Architecture: an Overview
	Author:	R. Braden, D. Clark, S. Shenker.
	Doc. no.:	RFC 1633
	Date:	June 1994
[RFC1883]	Title:	Internet Protocol; Version 6 (IPv6)
	Author:	S. Deering, R. Hinden
	Doc. no.:	RFC 1883
	Date:	December 1995
[RFC2205]	Title:	Resource ReSerVation Protocol (RSVP); Version 1 Functional Specification
	Author:	R. Braden, L. Zhang, S. Berson, S. Herzog, S. Jamin
	Doc. no.:	RFC 2205
	Date:	September 1997
[RFC2212]	Title:	Specification of Guaranteed Quality of Service
	Author:	J. Wroclawski
	Doc. no.:	RFC 2212

	Date:	September 1997
[RFC2215]	Title:	General Characterization Parameters for Integrated Service Network Elements
	Author:	S. Shenker, J. Wroclawski
	Doc. no.:	RFC 2215
	Date:	September 1997
[RFC2328]	Title:	OSPF Version 2
	Author:	J. Moy
	Doc. no.:	RFC 2328
	Date:	April 1998
[RFC2370]	Title:	The OSPF Opaque LSA Option
	Author:	R. Coltun
	Doc. no.:	RFC 2370
	Date:	July 1998
[RFC2386]	Title:	A Framework for QoS-based Routing in the Internet
	Author:	E. Crawley, R. Nair, B. Rajagopalan, H. Sandick
	Doc. no.:	RFC 2386
	Date:	August 1998
[RFC2475]	Title:	An Architecture for Differentiated Services
	Author:	S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, W. Weiss
	Doc. no.:	RFC 2475
	Date:	December 1998
[STALL98]	Title:	High-speed networks; TCP/IP and ATM design principles
	Author:	W. Stallings
	ISBN:	0-13-525965-7
	Date:	1998
[SZABO98]	Title:	QoS support in the Internet
	Author:	R. Szabo
	Doc. no.:	http://switchlab.ericsson.se/documents/routing.html
	Date:	1998-02-10
[VISWA98]	Title:	Evolution of Multiprotocol Label Switching
	Author:	A. Viswanathan, N. Feldman, Z. Wang, R. Callon
	Doc. no.:	IEEE Communications Magazine
	Date:	May 1998
[WHITE97]	Title:	RSVP and Integrated Services in the Internet: A Tutorial
	Author:	Paul. P. White
	Doc. no.:	IEEE Communications Magazine
	Date:	May 1997
[XIAO99]	Title:	Internet QoS: A Big Picture
	Author:	X. Xiao, L.M. Ni
	Doc. no.:	IEEE Network magazine

Date: March/April 1999

[ZAPPA97] Title: Alternate Path Routing and Pinning for Inter-domain Multicast Routing
Author: D. Zappala, D. Estrin, S. Shenker
Doc. no.: USC Computer Science Technical Report #97-655
Date: 1997

7.2 By category

The Internet Draft, RFCs, articles and reports are categorised into different areas to provide an overview of which document are available in which area. Furthermore, the type of document is listed between brackets.

7.2.1 RFCs

<u>QoS-based routing</u>

A Framework for QoS-based Routing in the Internet	[RFC2386]	(overview)

<u>OSPF</u>

OSPF Version 2	[RFC2328]	(specification)
The OSPF Opaque LSA Option	[RFC2370]	(specification)

<u>Differentiated Services</u>

An Architecture for Differentiated Services	[RFC2475]	(overview)

<u>Integrated Services</u>

Integrated Services in the Internet Architecture: an Overview	[RFC1633]	(overview)
General Characterization Parameters for Integrated Service Network Elements	[RFC2215]	(specification)
Resource ReSerVation Protocol (RSVP); Version 1 Functional Specification	[RFC2205]	(specification)
Specification of Guaranteed Quality of Service	[RFC2212]	(specification)

<u>Other</u>

Internet Protocol; Version 6 (IPv6)	[RFC1883]	(specification)

7.2.2 Internet Drafts

<u>OSPF</u>

An Overlay Model for Constraint-Based Routing	[IDOSPF99]	(protocol)
Multicast extensions to OSPF	[RFC1584]	(protocol)
OSPF Optimized Multipath (OSPF-OMP)	[IDOSPF99b]	(protocol)
Quality of Service Extensions to OSPF or Quality Of Service Path First	[IDOSPF97]	(protocol)
QoS Routing Mechanisms and OSPF Extensions	[IDOSPF98]	(protocol)

<u>RSVP</u>

Extended RSVP-Routing Interface	[IDRSVP97]	(protocol)
RSRR: A Routing Interface For RSVP	[IDRSVP98]	(protocol)
Setting up Reservations on Explicit Paths using RSVP	[IDRSVP97b]	(protocol)

<u>MPLS</u>

A Framework for Multiprotocol Label Switching	[IDMPLS97]	(overview)

<u>Routing</u>

Routing of Multimedia Connections when Interworking with PSTN, ATM and IP Networks	[IDROUTE98]	(overview)

7.2.3 ATM forum

Private Network-Network Interface Specification	[AF96c]	(specification)

7.2.4 Reports and articles

<u>Multicast</u>		
Alternate Path Routing and Pinning for Inter-domain Multicast Routing	[ZAPPA97]	(protocol)
Multi-Path Routing combined with Resource Reservation	[CIDON97]	(protocol)
<u>Measurements</u>		
End-to-end Routing Behavior in the Internet	[PAXSON97]	(measurements)
Implementation and Performance Measurements of QoS Routing Extensions	[APOST99]	(measurements)
<u>OSPF</u>		
Improving QoS Routing Performance Under Inaccurate Link State Information	[APOST99b]	(evaluation)
On Reducing the Processing Cost of On-Demand QoS Path Computation	[APOST98b]	(evaluation)
Quality of Service Based Routing: A Performance Perspective	[APOST98]	(evaluation)
<u>QoS-based routing</u>		
On Path Selection for Traffic with Bandwidth Guarantees	[MA97b]	(algorithm)
Quality-of-Service Routing for Traffic with Performance Guarantees	[MA97]	(evaluation)
QoS-based Routing in Networks with Inaccurate Information: Theory and Algorithms	[GUERIN97]	(evaluation)
QoS support in the Internet	[SZABO98]	(overview)
Routing Traffic with Quality-of-Service Guarantees in Integrated Service Networks	[MA98]	(evaluation)
Routing with End to End QoS Guarantees in Broadband Networks	[ORDA99]	(evaluation)
Supporting Dynamic Inter-Class Resource Sharing: A Multi-Class QoS Routing Algorithm	[MA99]	(evaluation)
<u>Other</u>		
Internet QoS: A Big Picture	[XIAO99]	(overview)
Wide Area Traffic: The Failure of Poisson Modeling	[PAXSON95]	(measurement)
RSVP and Integrated Services in the Internet: A Tutorial	[WHITE97]	(overview)

8 Appendix A Patents

The results of a search for patents in the area of QoS-based routing are presented. First the patent database is explained shortly, next the search results are presented.

8.1 Introduction

The Ericsson Patent database is an easily accessible reference to current patent literature based upon the Derwent Information World Patent Index database (DWPI) [PATENT99].

The database contains abstracts of published patents and patent applications from 70 countries and patenting authorities, covering the period from 1975 to the present day. These patent abstracts describe physics, electrical engineering and electronic engineering. These categories ensure that all telecommunications technologies will be included.

Each patent abstract makes reference to an engineering drawing or image, which, if available, is displayed in line with the text of the document.

There are currently about 3.5 million records in the database, increasing by 5 000-10 000 new documents each week. This makes the database arguably one of the largest single intranet databases in Europe at the present time. The total volume of text and images is approximately 22GB (as of Jan 1998).

8.2 Search results

Several text based searches have been carried out to evaluate the number of patents w.r.t. QoS-based routing. The results have been screened w.r.t. to their applicability to QoS-based routing.

No search results, relevant to QoS-based routing, have been found with a text search with keywords: OSPF, RIP, BGP or IDRP.

The following query provided 145 results: ROUTING **AND ((**QOS **OR** IP **OR** INTERNET**) OR (**OF **AND** SERVICE **AND** QUALITY**))**.

Time- and state-dependent PSTN routing (see Section 3.2):

Document	PAN 98-121105
Derwent Title	Database central routing device for incoming access requests compares time of day, peak and off-peak hours, and holiday information in order to select suitable access provider
Patentee details	WEBTV NETWORKS INC;(WEBT)
Abstract	The device comprises a central server and a user identifier which detects a user requesting access as a customer. An algorithm generator produces algorithms related to phone numbers of data base access providers. The central server downloads an algorithm on the basis of user criteria. A user identifier is an automatic phone number identifier or a log-in routine of a modem pool. The criteria includes time of day, date, holidays, busy signals, accumulated connect time, quality of access, reliability, performance of points of presence, peak times and geographical location.
Use Advantage	For use with several Internet access providers. Improved network transaction performance due to considering peak times. Considers costs of different methods of access.
Title Terms	database central route device incoming access request compare time day peak peak hour holiday information order select suit access

QoS metrics (see Section 4.3):

Document	PAN 99-112753

Derwent Title	Routing method in communication network such as ATM network, internet involves deciding QOS class based on routing demand for determining single scale minimum path, by transforming metric and attribute using information received from one switching system
Patentee details	NIPPON TELEGRAPH & TELEPHONE CORP;(NITE)
Abstract	NOVELTY - Transformation of a metric and attribute is performed to a single quality scale for every QOS class, by a switching system in predetermined manner, when a metric and attribute connection information is informed from another switching system, based on necessity. The transformation is performed to decide the minimum path corresponding to the metric more than two or the final desired value of the attribute. The QOS class is decided corresponding to the terminal number based on a routing demand, when a call is received.
Use Advantage	In ATM network, internet for multimedia applications. Selects optimum route satisfying several metrics and attributes. DESCRIPTION OF DRAWING(S) - The figure shows the block diagram of communication network in which the routing method is employed.
Title Terms	route method communicate network atm network decide class based route demand determine single scale minimum path transform metric attribute information receive one switch system

PSTN routing (see Section 3.2):

Document	PAN 98-584300
Derwent Title	Call routing method on outgoing trunks involves segregating calls that are likely to cause congestion by using special prefix code and forming alternate virtual trunk being percentage of total trunk lines out from exchange
Patentee details	BELL CANADA; STENTOR RESOURCE CENT INC;(BELL ; STEN)
Abstract	The method involves a calling party (25), who wishes to gain access to a special type of call use, using a prefix code (e.g. 1+) when dialling. The local exchange service switching point (SSP) (27) segregates calls containing this prefix code and selects an alternate path for the call between SSP and access tandem (28). The selection is made based on the type of prefix code; the code being looked up in routing table (30) and the specified alternate path (virtual trunk) (26) then set up. The virtual path may dynamically alter based on time of day, usage rate, and trunk capacity. Segregation can also be automatically done using automatic route selection provided by an advanced intelligent network.
Use Advantage	For subscribers accessing special services, e. g. promotional long distance call rates and access to internet. Minimises call congestion experienced by residential, business and regular long distance subscribers, whilst still offering access to special calls.
Title Terms	call route method outgoing trunk segregate call cause congested special prefix code forming alternate virtual trunk percentage total trunk line exchange

QoS-based routing:

Document	PAN 98-452892
Derwent Title	Internet protocol packet routing control method in LAN involves indicating degradation level of packet among communication routers to server based on packet loss algorithm
Patentee details	NIPPON TELEGRAPH & TELEPHONE CORP;(NITE)
Abstract	The method involves communicating a data packet to a receiver, based on the specified address in the IP packet header through several routers (10-16). The packet transmission loss characteristic in each router is monitored. The packet degradation level is compared with a predefined level in each router. The information regarding the packet degradation level and the degradation packet address is transmitted to other routers. Based on a predefined packet, loss algorithm, the packet loss along the communication channel is indicated to the server.
Use Advantage	Enables effective utilisation of network resource.
Title Terms	protocol packet route control method lan indicate degrade level packet communicate router serve based packet loss algorithm

QoS-based routing:

Document	PAN 97-379145
Derwent Title	QOS routing apparatus for communication network has path selector which connects selected path if quality of service is satisfied, otherwise, optimum path which will satisfy QOS will be selected
Patentee details	NEC CORP;(NIDE)
Abstract	The routing apparatus has a link state routing protocol unit (140) which stores the topology and the link communication quality in a link state database (141). A signalling unit (145) performs a connection establishment procedure upon receiving a connection establishment message. A candidate path is chosen based on data stored in the link state database through a candidate path selector (142). A quality of service decision unit (143) determines whether the QOS in the selected path is satisfied. A path selector (144) connects the selected path if the QOS is satisfied, otherwise, another path is selected.
Use Advantage	Reduces connection establishment delay by reducing amount of calculations needed for choosing optimal path.
Title Terms	route apparatus communicate network path select connect select path quality service satisfy optimum path satisfy select qos

QoS-based routing:

Document 133	PAN 95-403710
Derwent Title	Preferential resource constraint consolidation combining requested and acceptable resource constraint sets for efficient solution to problem of consolidating preferential resource constraints
Patentee details	MOTOROLA INC;(MOTI)

Abstract	The device for providing consolidation of preferential resource quality of service constraints specified by a source user and a destination user in a communication network has two quality of service translators receiving preferential resource constraints input from the source and destination users respectively. The translators translate input QOS requirements, for each of the source and destination users, into requested and acceptable resource constraint sets. A QOS consolidator consolidates the resource constraint sets in accordance with a predetermined merge scheme to provide appropriate constraints for path selection and connection establishment. A routing function unit selects an acceptable path for connection establishment between the source and destination users. A network resource allocator allocates resources along the selected acceptable path for connection establishment that supports information transfer between the source user and the destination user.
Use Advantage	Consolidates resource constraints to resolve potential differences in resource performances of end users.
Title Terms	prefer resource constrain consolidate combination request accept resource constrain set efficiency solution problem consolidate prefer resource constrain

QoS-based routing:

Document 134	PAN 95-373454
Derwent Title	Quality of service manager for source routing multimedia packet LAN uses hierarchical database in management system to allocate resources to stations with network in response to requests
Patentee details	INT BUSINESS MACHINES CORP;(IBMC)
Abstract	The local area network has several user stations interconnected by transmitters. A data path is selected between two of the stations. A quality of service management unit is connected by the transmitters to the user stations for ensuring that each path selected meets pre-defined quality of service parameters. The QOS management unit has a memory storing a number of parameters for each transmission component. A comparator, responsive to a request for a multimedia path between two user stations, compares each requested path component parameter to the corresponding stored parameter for the requested path components. A path is allocated when the stored parameters, for each component of a requested path, equal or exceed those requested and denied when the stored parameters for at least one component of a requested path is less than the corresponding requested parameter

The following query provided 49 results: QoS

QoS-based routing:

Document	PAN 99-134413
Derwent Title	Route selection method e.g. for communication route selecting route satisfying several Quality of Service conditions, out of several routes connecting start point to end point via at least one node
Patentee details	KOKUSAI DENSHIN DENWA CO LTD;(KOKU)
Abstract	NOVELTY - The method involves selecting a route satisfying several Quality of Service (QoS) conditions, out of several routes connecting a start point to an end point via at least one node. DETAILED DESCRIPTION - The method sets conditions to be satisfied for several QoSs, respectively and inputs an error range permissible for a cost of an unknown optimum route having a minimum cost among routes satisfying all QoS conditions. A cost searching range containing at least the minimum cost is set and a determination is made whether the cost searching range has been narrowed to a searching possible range, the range is a function of the error range. In response to judgement that the cost searching range has not been narrowed to the searching possible range, several QoSs of routes leading from the start point to respective nodes at each cost are derived, in order of cost, beginning with a lowest cost within a current cost searching range, on the basis of QoSs of nodes having QoSs already derived. The cost searching range is narrowed when a route leading from the start point to the end point and satisfying all of the QoS conditions is found, on the basis of it cost. The cost sea rching range narrowed to the searching possible range an optimum route is searched.
Use Advantage	For communication route. Optimum route held in predetermined error range can be selected without depending upon operator's knowledge or experience concerning Quality of Service conditions. DESCRIPTION OF DRAWING(S) - The figure shows a schematic flow chart of a route selection method.
Title Terms	route select method communicate route select route satisfy quality service condition route connect start point end point one node

The following query provided 665 results: ROUTING **AND** SELECTION

QoS-based routing:

Document	PAN 99-073225
Derwent Title	Processor assisted routing table generation method for telecommunications network using different combinations of partial routing tables for providing several overall routing tables, subjected to quality evaluation, for selection of new overall routing table
Patentee details	SIEMENS AG;(SIEI)
Abstract	The routing table generation method involves providing the routing table, using partial routing tables which are combined to provide several overall routing tables. Each of these is evaluated for assignment of a corresponding quality value, used for selection of the new overall routing table. Several of partial routing tables can be provided for each source node/ target node

	pair, marked by respective identification values. The partial routing tables with the same identification values are combined to provide each overall routing table.
Use Advantage	For optimum call connection routing in telecommunications network. Rapid communication with minimum data loss.
Title Terms	processor assist route table generate method telecommunication network combination route table overall route table subject quality evaluate select new overall route table

QoS-based routing:

Document	PAN 94-049141
Derwent Title	Call by call source routing method attempting to connect on constrained routing basis and then falling back to look around first form of pre-emption if optimal routing fails
Patentee details	MOTOROLA INC; CODEX CORP;(MOTI ; RENI)
Abstract	The system and method for handling call-by-call source routing includes a rule based fallback approach. The integrated traffic system allocates priorities to calls as they arrive and in various connection states. Initially, an attempt to place the call by selecting (202) a constrained routing is made, including deciding on an optimal feasible path (204) and attempting to connect the call. If this fails the fall-back strategy is applied (206) alternative routing is obtained by using alternative paths based on sequence of constrained routing by predetermined or real-time selection.
Use Advantage	Identifying alternative acceptable paths for integrated traffic types. Provides flexible platform to perform routing in integrated traffic system with quality-of-service limits.
Title Terms	call call source route method attempt connect constrain route basis fall back first form pre optimum route fail

The following query provided 16 results: CONSTRAINT **AND** ROUTING

QoS-based routing:

Document	PAN 96-354173
Derwent Title	Routing path selection method using network management node in telecommunication system calculating path loading vectors, constraint vector, and switch congestion vector and comparing to yield potential intermediate switch candidates having lowest available trunk traffic loading
Patentee details	AT & T IPM CORP; AT & T CORP;(AMTT ; AMTT)
Abstract	In a telecommunication network having a number of switches interconnected by groups of communication channels, each switch periodically generating traffic data indicative of the volume of calls on the respective groups of communication channels, a method for automatically selecting a routing path for calls from an originating switch via an intermediate switch to a destination switch, involves receiving and storing at a network management node the traffic data from a set of the switches. The latter contains certain switches which are not homogeneous and do not share the same operating architecture. At least a first vector is generated at the network management node based on the stored traffic data, the first vector contg. a first element for each group of communication channels associated with one switch in the set of switches, the first elements defining if a predetermined traffic level exists on the corresp. groups of communication channels associated with the one switch. A path constraint vector is generated at the network management node based on predetermined routing factors stored at the network management node, the path constraint vector contg. an element for each switch. Each of the path constraint vector elements defines if a corresp. switch is available to serve as the intermediate switch. The first vector elements and corresp. path constraint vector elements are compared to produce a resulting selection vector with an element corresp. to each switch in the set of switches. The constraint vector elements determines whether a corresp. switch is a candidate for selection as the intermediate switch. One of the candidate switches is selected as the intermediate switch through which calls from the originating switch pass to the destination switch.
Use Advantage	Trunk groups with increasing levels of traffic and switches with increasing levels of congestion are incrementally tested in order to yield potential intermediate switch candidates and calls are distributed to lightest loaded trunks and switches.
Title Terms	route path select method network management node telecommunication system calculate path load vector constrain vector switch congested vector compare yield potential intermediate switch candidate low available trunk traffic load