University of Würzburg
Institute of Computer Science
Research Report Series

Possibilities for QoS in existing Internet Routing Protocols

Richard Harris[1] and Stefan Köhler[2]

Report No. 228 May 1999

[1] RMIT
City Campus
GPO Box 2476V
Melbourne 3001 Victoria Australia
richard@catt.rmit.edu.au

[2] University of Würzburg
Department of Computer Science
Lehrstuhl für Informatik III
Am Hubland, D-97074 Würzburg, Germany
koehler@informatik.uni-wuerzburg.de

Possibilities for QoS in existing Internet Routing Protocols

Richard Harris
RMIT
City Campus
GPO Box 2476V
Melbourne 3001 Victoria Australia
richard@catt.rmit.edu.au

Stefan Köhler
University of Würzburg
Department of Computer Science
Lehrstuhl für Informatik III
Am Hubland, D-97074 Würzburg, Germany
koehler@informatik.uni-wuerzburg.de

Abstract

In classical circuit switching networks Quality of Service is an essential aspect which influence the entire network design process, including routing consideration. Due to the efforts to introduce Quality of Service (QoS) into IP-based networks, design routing to support QoS becomes increasingly important, e.g. in such concepts like integrated service and differentiated service IP networks. In this we give an overview of the possibility to introduce QoS in existing routing schemes and give an outlook to new QoS routing protocols currently discussed in the internet community such as the IETF. First of all we give a survey of different routing algorithms used in unicast Interior Gateway Protocols and Exterior Gateway Protocols. We describe in detail (E)IGRP a proprietary CISCO routing protocol, followed by a short overview of the different aspects for QoS routing protocols in the Internet.

1 Introduction

Recently, we have seen the merging of circuit switched network services and data services into integrated digital networks. The characteristics of the various types of service when they are combined onto one type of integrated network must be taken into account. Furthermore, we find that these characteristics are often incompatible and that complex issues relating to Quality of Service (QoS) have arisen. In delivering communication services to customers these QoS issues are directly affected by the routing technologies that are employed. Over the past few decades, telephony networks and packet switched data networks have evolved quite separately from each other and they each have employed quite different routing strategies to deliver services to their customers.

We consider the problem of introducing optimal routing methods for IP packet traffic. IP datagrams are currently used to transmit many different types of network services. The rapid explosion in the use of the Internet for Web browsing, telephony and video services, as well as more traditional services such as telnet and ftp, has resulted in a massive increase in traffic load. With this increase in load, there has been a corresponding significant increase in congestion due to the lack of network resources. Efficient routing of IP packets is becoming a crucial issue both from the point of view of the providers and the users of the network. The many different services can have quite different quality of service requirements and consequently, the routing strategies to be employed in the network will need to distinguish in some way between the different types of traffic. Over the past few decades, telephony networks have employed both fixed and dynamic routing strategies and there is now a significant amount of literature that underpins the application of these strategies [6, 1, 2]. By way of contrast, packet based networks have evolved very rapidly and routing strategies have focused on delivering data to the destination without the need to consider timing issues. Although most routing algorithms that have been developed for packet networks have identified the essential parameters that impact upon the timely and reliable delivery of packets to their destination, the configuration of routers has been undertaken rather arbitrarily and (possibly) in a non-optimal manner.

In this report, we shall discuss some of the following issues:

- What is the current methodology used for routing IP traffic?
- What is the methodology for passing routing information around the network?
- What is the impact of different routing strategies on the quality of service for various service types.

Messages sent over a communication network can be classified into three broad headings, viz:

- Unicast: The message is transmitted from the source to a single (specific) destination. Such a transmission is often referred to as "point- to-point" communication.
- Broadcast: The message is broadcast to all possible destinations in the network.
- Multicast: The message is broadcast to a specified subset of customers (destinations) in the network.

Unicast communication has been the main mode of communication in both packet and circuit switched networks; however, more recently, the growing use of conference calls, resource discovery and multimedia conferences has generated major interest in the need for routing strategies that result in packets/messages that are destined for a group address (multicast communication).

2 Current IP Unicast Routing Strategies

In the sequel, we shall briefly review some of the routing methods employed in packet switching networks for unicast communications.

2.1 Classification of Mechanisms

Up until the present time, there have been a number of different unicast IP routing strategies employed by router manufacturers such as CISCO, Wellfleet, 3Com, etc (Fig. 1). They can be broadly classified as:

- Interior Gateway Protocols (IGP) and
- Exterior Gateway Protocols (EGP).

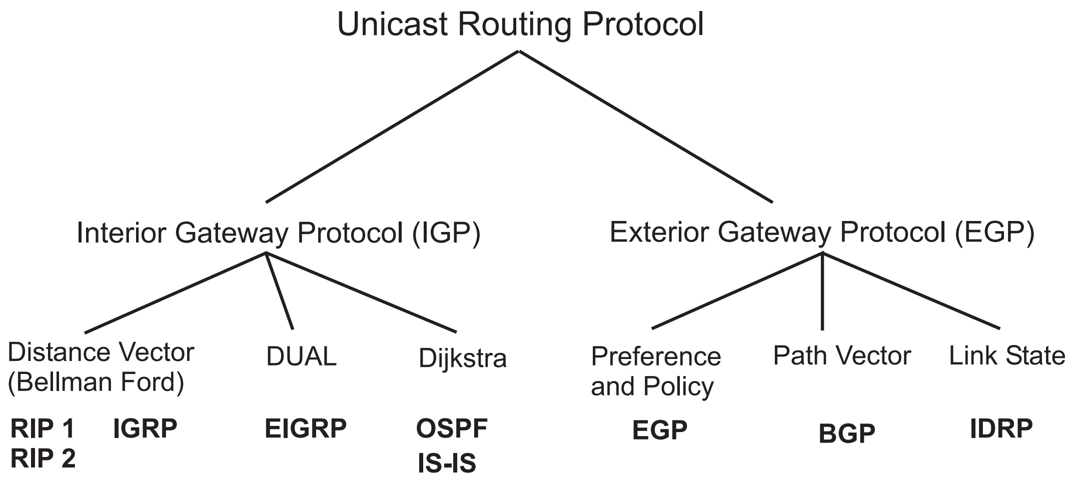


Figure 1: Overview of Unicast Routing Protocols

The IETF defines these protocols as interior gateway protocols if they are used for "routing networks that are under a common network administration". This common administration is frequently referred to as an *Autonomous System (AS)*. The interior routing protocols supported by CISCO include the following:

- Routing Information Protocol (RIP) – This is a commonly used Internet Gateway Routing Protocol created for use in small, homogeneous networks. It is a standard distance-vector routing protocol[1], which was initially designed as a component of the networking code for the BSD release of UNIX.
- Internet Gateway Routing Protocol (IGRP) – A CISCO proprietary routing protocol, which was developed 1986 as an extension of RIP.
- Enhanced Internet Gateway Routing Protocol (EIGRP) – An extension to the CISCO IGRP routing protocol.

[1] E.g. the University of Würzburg uses RIP for routing its campus network.

- Open Shortest Path First (OSPF) – Open Shortest Path First is a shortest path routing protocol that was specifically designed for IP networks. Details of the first version of this specification are described the Internet Engineering Task Force(IETF) document [9]. The latest version of OSPF is defined in [10] and the version number is 2.

Exterior gateway protocols as defined by CISCO "exchange routing information between networks that do not share a common administration." The exterior gateway protocols supported by CISCO include:

- Exterior Gateway Protocol (EGP)
- Border Gateway Protocol (BGP)
- Inter-Domain Routing Protocol (IDRP)

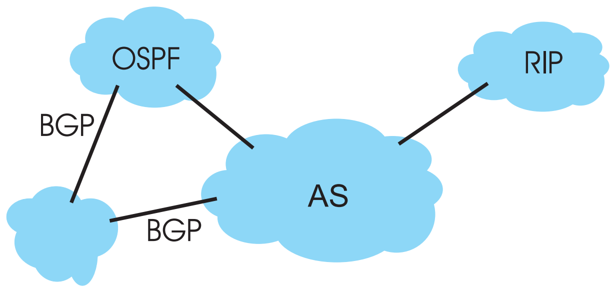


Figure 2: IGP (e.g. RIP, OSPF) protocols are used to route in an Autonomous System (AS). EGP (e.g. BGP) are used to route between AS.

Fig. 2 shows that these protocols operate across/between Autonomous Systems. Each of these protocols needs to be configured for use in its respective network environment. CISCO publishes details for their customers on how to configure their routers for these protocols, although they do not provide guidance on how to choose these settings. It should also be noted that the EGP protocol has all but vanished from the scene and thus we only describe the Border Gateway Protocol, which is up to version 4 at the time of writing this paper. In the next few sections, we shall provide a brief summary of how these protocols operate and the nature of the parameters that are required for their operation.

2.2 IP Routing Protocols and Parameters

Over the years, a number of parameters have been identified for use in the selection of routes for IP packet traffic. These typically include the following:

- Current link delay

- Load utilization
- Link reliability
- Available bandwidth
- Physical distances between nodes – shortest path
- Hop counts
- Split route percentages

Other parameters can also be used for route selection. In fact, it is quite common to combine these parameters and establish some form of weighted-metric (see section 2.5.4 for an example).

2.3 Performance Measures of Routing Algorithms

To do performance measurement and to compare the different routing algorithms it is necessary to identify and characterize the following performance parameter:

- Throughput – Average rate of successful packet transmissions per unit time.
- Average packet delay – Average time taken for packets to traverse the network from source to destination.
- Message complexity – Average number of messages sent for algorithm completion.

2.4 Comments on Current Schemes

The main objectives of IP routing schemes (or any packet switching based network) are to:

- maximize throughput
- minimize delay
- minimize the number of hierarchical levels (implied by the second rule)

Mostly it is not possible to optimize throughput and delay. To optimize throughput the traffic must distribute evenly over the network. On the other side, to optimize delay the shortest path has to be chosen.

After talking to administrators who manage network routers on a daily basis, a pattern is emerging that indicates a general lack of understanding of the parameters used for the routing process and their impact upon performance. In many cases, administrators simply accept the factory default settings in the expectation that this will be adequate for normal operation of the router. However, the number and range of parameter settings available is quite significant and most administrators would not try and adjust them - if their system is seen to be "working". Studies of IP routing performance have been undertaken in a few cases and the results seem to suggest that routing "anarchy" tends to prevail through the multitude of different routing algorithms that are available.

2.5 Interior Routing Protocols (IRP)

2.5.1 Distance-Vector-Routing Protocols

Distance-Vector (DV) Routing protocols require that each router simply informs its neighbours of its routing table contents. For each network path, the receiving routers pick the neighbour advertising the lowest cost, then add this entry into their routing table for readvertisement.

RIP is a common DV routing protocol. The well-known weakness of DV routing protocols is the appearance of routing loops. The phenomena is also know as the counting to infinity problem. Consider the configuration of Figure 3, with all link costs equal to one.

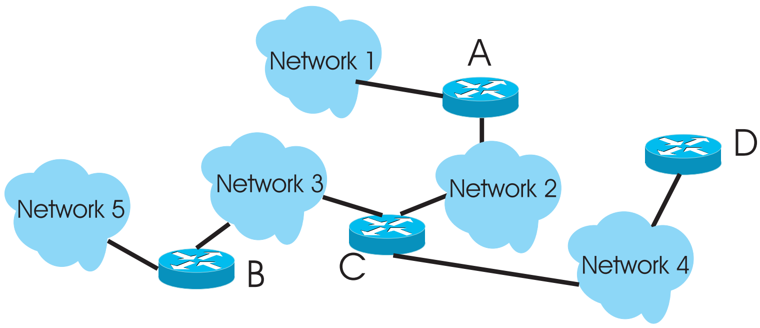


Figure 3: Counting to infinity problem [17]

Router C maintains a distance of 2 to network 5, with the next hop B, while Router A and D maintain a distance of 3 to network 5, with the next hop to C. If now router B fails, the following could happen:

1. Router C notices that network 5 is no longer directly reachable and sets its distance value to 4, due to the report from A and D.
2. At the next routing update, C advertises this information to A and D.
3. Router A and D receive this increased distance information from router C and increment their reachability information to network 5 to a distance of 5.

4. Router C receives the new distance count of 5 and calculates 6 (= 5+1) as the actual distance to network 5.

This procedure continues until the distance value reaches a predefined infinity threshold, e.g. 16 for RIP. If infinity is reached, the router recognizes that the target network is no longer reachable. In RIP with a reporting interval of 30 seconds, this type of condition can take several minutes.

The intermediate state of the network is very instable during the count to infinity: packets loop, so links will be congested, and routing packets themselves can be lost due to congestion. This is the reason why the maximum path cost is generally set to a small value in DV routing protocols. Thus, the metric range is not very broad.

To avoid these loops DV algorithms are enhanced to include triggered updates, split horizon, poison reverse and hold-down. We shall discuss these enhancements in the following subsection.

2.5.2 Loop Avoidance in Distance Vector Routing Algorithms

- Triggered updates
  In response to some changes in the routing table, the new routing table is sent immediately.
- Split horizon
  It is not useful to send information back on a route from which the information came. In Fig. 4 Router 1 (R1) knows from its initialisation that there is a route to network A. There is no reason for Router 2 (R2) to include this route in its return update. The split horizon rule says that R2 should delete this route from any updates that it sends to R1.

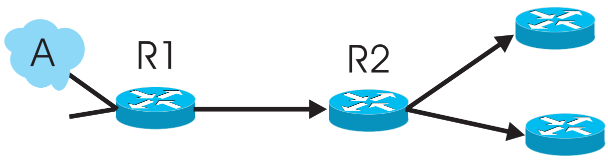


Figure 4: split horizon rule

- Hold-down
  Hold-downs tell routers to hold down any changes that might affect recently removed routes for some period of time. Hold-downs are used to prevent regular update messages from inappropriately reinstalling a route that has gone bad.
- Poison reverse
  Poison reverse is an improvement on split horizon. The idea is that increases in routing metrics generally indicate routing loops. To prevent larger routing loops, poison reverse updates are sent to remove the route and place it in hold-down.

2.5.3 Routing Information Protocol (RIP)

RIP was designed for a reasonably homogeneous, small- or moderate-sized network. However, in larger, more complicated internetworks, RIP has several drawbacks. First, RIP's hop count limit is 16, so destinations may not be more than 15 hops distant. Second, the protocol cannot choose routes based on real- time parameters such as delay or load. The metric is the hop-count to the host whose IP address is quoted. Because of the small metric range $[0, 16]$ there is no opportunity to prefer a T1 link over a 64KBit/s link. RIP information is transmitted using UDP/IP[2] on port number 520. RIP broadcasts a complete list of networks it can reach every 30 seconds to its neighbors. Depending on the lengths of theses lists, bandwidth usage can become prohibitive on slow links. RIP suffers from very slow convergence in the face of topology changes because routers are not under any obligation to identify failed links like OSPF with Hello packets and, more importantly, their consequences and propagate the facts to other routers (count to infinity problem). RIP itself has no security features. RIP and the Distance Vector, or Bellman-Ford algorithm is specified in [7]. The only QoS-aspect RIP satisfy is well known in circuit switched networks. There, one tries to minimize the use of resources and that's exactly what RIP achieves with the hop-count algorithm.

There is a newer version, known as RIP-2 [8]. Due to the poor distribution of RIP-2 we list only the improvements:

- Variable length subnet masks (VLSM) support
- Route summarisation
- Classless InterDomain Routing (CIDR)
- Multicast routing updates (address 224.0.0.9), although all RIP-2 routers can be configured to revert to broadcast in order to inter-operate with RIP-1 routers
- Authenticated updates using MD5

To be compatible with RIP-1, RIP-2 has the same metric range as RIP-1. Consequently RIP-2 has inherited most of the disadvantages from RIP-1. RIP has also been modified to carry IPv6 addresses, resulting in the RIPng routing protocol for IPv6.

2.5.4 Interior Gateway Routing Protocol (IGRP) and Enhanced IGRP

IGRP is an intra-domain distance vector routing protocol developed in the mid-1980s. IGRP uses a combination (vector) of metrics. Internetwork delay, bandwidth, reliability, Maximum Transfer Unit (MTU), and load are intended the routing decision. IGRP permits multi-path routing. Multiple equal-bandwidth paths may run a single stream of traffic in round-robin fashion, with automatic

[2] no transmission/error control

switchover to the second line if one line goes down. With Software Release 9.21 of CISCO's IOS, CISCO introduced an enhanced version of IGRP that combines the advantages of link state protocols with the advantages of distance vector protocols.

Enhanced IGRP Features

EIGRP uses the same distance information as IGRP. However, the convergence properties and the operating efficiency of EIGRP have improved significantly over IGRP, due to that EIGRP incorporates the Diffusing Update Algorithm (DUAL) [12].

- Fast convergence. The DUAL (Diffusing Update Algorithm) is free of transient loops, however it has to introduce transient unreachabilities.
- Partial updates. Enhanced IGRP sends incremental updates when the state of a destination changes, instead of sending the entire contents of the routing table.
- Neighbour discovery with a simple "Hello" mechanism. The mechanism is protocol independent.
- Support of VLSM.
- Bounded updates - Enhanced IGRP does not make periodic updates. Instead, it sends partial updates only when the metric for a route changes. Propagation of partial updates is automatically bounded so that only those routers that need the information are updated.
- Multiple network layer support;
- Better Scaling;

Metric

The calculation of the metric and default metric value for IGRP and Enhanced IGRP are nearly the same. The main difference is that IGRP uses a 24bit while EIGRP uses a 32bit field to store the values. (E)IGRP calculates the metric M using the following equation:

$$M = \left[K_1 \cdot \text{bandw} + \frac{K_2 \cdot \text{bandw}}{256 - \text{load}} + K_3 \cdot \text{delay} \right] \cdot \frac{K_5}{\text{reliability} + K_4} \qquad \text{if } K_4 \neq 0 \text{ and } K_5 \neq 0 \tag{1}$$

else

$$M = K_1 \cdot \text{bandw} + \frac{K_2 \cdot \text{bandw}}{256 - \text{load}} + K_3 \cdot \text{delay} \tag{2}$$

where bandw $= \frac{10{,}000{,}000\text{bit}}{\min_{i=1\ldots n}(\mathrm{B}_i)}$ with B_i as the unload bandwidth of the link in kilobits per second and delay $= \sum_{i=1..n} \text{delay}_i$ were $1..n$ denotes the path.

The unloaded bandwidth of each network and the topological delay of each network are defined in the router configuration. Only the MTU, the reliability and the load of the channel are estimated through measurement.

	Meaning	**Range**
B	B is the bandwidth of the narrowest bandwidth segment of the path in kilobits per second. Bandwidth is deduced from the interface type.	1200bps - 10Gbps $[0; (2^{32}-1)]$ EIGRP $[0; (2^{24}-1)]$ IGRP
load	Effective bandwidth of the route. Load is dynamically computed as a 5 minute exponential weighted moving average that is updated every 5 sec.	[1; 255] A load of 255 indicates a completely saturated link
delay	Topological delay time. Topological delay time is the amount of time it takes to get to the destination along that path, assuming an unloaded network. In practice, a standard delay (default value T1-link) is used. The delay parameter in EIGRP is stored in a 32-bit field, in increments of 39.1 nanoseconds. This results in a range of 39,1 ns to 4,294,967,040 ns.	$[0; (2^{32}-1)]$ EIGRP $[0; (2^{24}-1)]$ IGRP
reliability	Reliability of the path. Reliability indicates the current error rate. It is the ratio of expected packets to the actual packet that arrive at the destination undamaged. Reliability is dynamically computed as a moving weighted average over five seconds.	[0; 255] 255 is 100 percent reliability or a perfectly stable link

Two additional data elements are included in routing updates:
The hop count[3] and Maximum Transmission Unit (in bytes), although are neither currently used in the calculation.

The default constant values are $K_1 = K_3 = 1$ and $K_2 = K_4 = K_5 = 0$, so the metric is reduced to:

$$M = \text{bandw} + \text{delay} \tag{3}$$

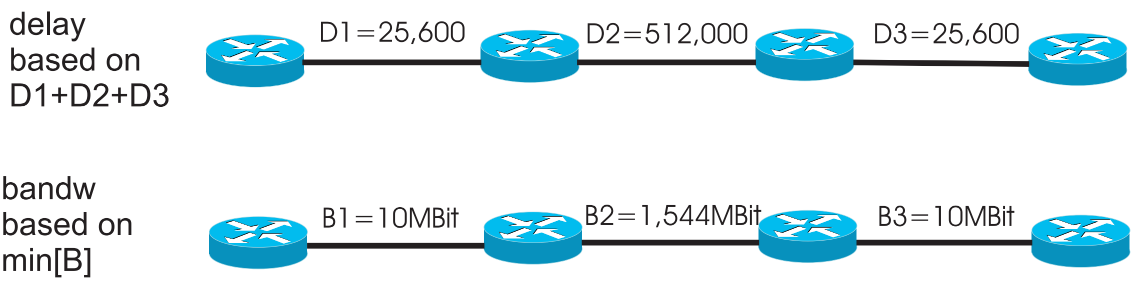
The minimum value of the above metric is select to decide the best path for routing of the packets.

[3] hop count: the largest possible value is 254. The hop counter in EIGRP is large enough to support thousands of hops, the only barrier to expanding the network is the hop counter of the transport layer.

	delay		bandwidth	
	IGRP	EIGRP	IGRP	EIGRP
Satellite (500Mbit)	200,000 (2s)	5,120 (2s)	20	5,120
Ethernet (10Mbit)	100 (1ms)	25,600 (1ms)	1,000	256,000
T1 (1,544Mbit)	**2,100 (21ms)**	**512,000 (20ms)**	**6,476**	**1,657,856**
64kbit	2,000 (20ms)	512,000 (20ms)	156,250	40,000,000
56kbit	2,000 (20ms)	512,000 (20ms)	178,571	45,714,176
10kbit	2,000 (20ms)	512,000 (20ms)	1,000,000	256,000,000
1kbit	2,000 (20ms)	512,000 (20ms)	1,000,0000	2,560,000,000

Table 1: Delay and Bandwidth [24]. T1 is the default setting of a CISCO-router. (Enhanced IGRP values = 256 · old values(IGRP))

With the values in Table 1 we can give an example for the default metric calculation:



$$M = \text{bandw} + \text{delay}$$

$$M = 256 \cdot \frac{10,000,000}{1,544} + (25,600 + 512,000 + 25,600) = 2,221,056$$

If the minimum bandwidth has high value (as in a core network), the delay is getting very important in the metric calculation. On the other hand, if the minimum bandwidth is small (e.g. a ISDN link 10,000,000/64 = 156,250) the delay can be neglected in most of the calculations.

The full Metric (1) is more complex. In the following paragraph we shall provide some comments on the different K-parameters.

- When $K_1 = 1$, it follows that we are trying to locate a path that will provide the largest bandwidth. Bandwidth means that you use the narrowest bandwidth on your path and you estimate the bandwidth out of your router configuration. This approach is similar to the DCR method for circuit switched routing. Note, however, that DCR routing included a random element in order to prevent the possibility that an exchange could be misreporting available capacity due to the presence of "killer trunks". In such a case, this misinformation could attract traffic to routers only to find that it would lose packets or in some way prevent correct delivery.

- When $K_3 = 1$, we are attempting to locate paths that minimize delay. Similar comments to those given above apply to this case also. Misreported information can lead to problems in sending the packets to their destination. Because you did not measure the delay, you only look in the configuration file of the router. If you only use the default value you get essentially a hop count metric, not a delay metric. In most packet switching networks, the objective is to minimize delays and so this particular parameter will figure in most routing algorithms in some form or another.

- When K_2 is set, this gives a combination of bandwidth and load. This is the first parameter which is measured. The parameter represents the real traffic situation on your network. As your network traffic is time dependent, you have a dynamic element in your calculation. In summary, paths with the minimum value of this option will be selected on the basis of available bandwidth, provided that the load is not too high - otherwise packets will be diverted to other paths. Because of the possible path oscillation, K_2 should be handled with care.

- If K_5 and K_4 are set, then the formula becomes even more complex. In this case, the term in brackets is scaled by the factor: $\frac{K_5}{\text{Reliability}+K_4}$, and this means that as the reliability increases, the overall metric will decrease - all other factors remaining equal.

In conclusion, the metric derived from the above formulae are not definitive and permit many degrees of freedom - even assuming that the bandwidth, load, reliability and delay are the only factors involved in determining an "optimal" routing strategy for IP packets.

Timing issues

In [18] S. Floyd and V. Jacobson studied the synchronization of periodic routing messages. They showed that the selection of the routing protocol timers is very important with respect to routing stability. They were able to show that synchronization can be avoided by the addition of randomization to the traffic sources and quantified how much randomization is necessary. In EIGRP it is possible to adjust the following timers in the configuration file of a router:

... time	Meaning	Default value
(broadcast) update	The interval (in seconds) between two routing updates. If there is a change, the routing table is sent immediately.	90 sec
invalid	The interval of time (in seconds) after which a route is declared invalid. A route becomes invalid when there is an absence of updates that refresh the route.	$3 \cdot \text{update}$
holddown	The interval of time (in seconds) during which routing information regarding better path is suppressed. It should be at least three times the value of update. A route enters into a hold-down state when an update packet is received that indicates the route is unreachable. The route is marked inaccessible and advertised as unreachable. However, the route is still used for forwarding packets. When holddown expires, routes to the same destination advertised by other sources are accepted and the route is no longer inaccessible.	$3 \cdot \text{update} + 10\text{s}$
flush	The amount of time (in seconds) that must pass before a route is removed from the routing table. The value should be greater than the sum of invalid and holddown time.	$7 \cdot \text{invalid}$
sleep	The amount of time for which routing updates will be postponed. This interval is in milliseconds and is used for postponing routing updates in the event of a flash update. The sleep value should be less than the update time. If the sleep time is greater than the update time, routing tables will become unsynchronized.	

Variance

In the event that two (or more) paths have the same value for their metric, the protocol specifies that the traffic should be split between these paths in proportion to their composite metrics. Thus, if two link speeds are 9.6kbps and 19.2kbps, but the metrics are the same then the packets will be sent in the proportion 1:1. The situation where the metrics are identical is rare in practice. However, the parameter known as “variance” and designated by the letter V has been proposed for certain types of routers. This value is used to provide a range of values for path metrics that can be regarded as “equal” by the router so that splitting of the traffic can be performed. To illustrate the approach, let us suppose that there are three outgoing routes from a router and the associated metrics have values 100, 120 and 250 respectively. Let us further assume that a user has selected a value V=2. Then this “variance” is applied in the following way. We select the minimum metric value in the usual way (=100) and consider the range of metric values 100 to $100 \cdot V = 200$ as being equal for the purpose of splitting the traffic. Thus, there are two paths in this range (the third value of 250 is outside of this range so it is not included in the split). Traffic is now split over the two paths with metrics 100 and 120 in proportion to their metric. Thus we see that we have the equivalent of circuit switching's concept of split route trunking in a packet switching environment! Furthermore, if a large value for V is specified, then all eligible routes could be used in the split. The value of the variance can be from 1 to 128. The default is 1. There are three main problems in using the parameter variance. The first danger is that with a large enough variance, paths become allowed that aren't just slower, but are actually "in the wrong direction". The second problem is more a technical one. Consider a UDP connection. If the underlying layer is, for example, Frame Relay, there is no mechanism which guarantees correct sequencing of packets. The third one is that it is easy to build up partial loops through ill-tuned multipath routing [16].

Type of Service (TOS)

One aspect of QoS routing is that different applications need different routing strategies. A real time application should have a guaranteed delay and bandwidth on a fixed path from source to destination. For ftp it is not necessary to satisfy such strong conditions.

Many authors suggest a classification of traffic types which could form the basis for quality of service guarantees in multi-service networks. Like in a recent paper [13] from Jim Roberts, these two classifications are stream and elastic traffic. He defines these types as follows [14]:

- Stream traffic entities are flows having an intrinsic duration and rate (which may be variable) whose time integrity must be (more or less) preserved by the network; such traffic is generated by applications like the telephone and interactive video services such as video conferencing.
- Elastic traffic entities are digital objects that must be transferred from one place to another; these objects might be files of alphanumeric data, texts or pictures, for example.

In EIGRP it is possible to define for every TOS parameter a suitable selection of the K-parameters. If an application uses this Type of Service field, we have taken the first step to QoS routing. The next step should be the implementation of a mechanism which reserves bandwidth like RSVP [21].

Link-State-Routing Protocols

Definition of Link-State (LS) Routing:
Each router sends routing information to **all** other routers. The information contains the exact value of its link cost to adjacent networks.

Open Shortest Path First (OSPF Version 2)

Open Shortest Path First is a routing protocol developed by the interior gateway protocol (IGP) working group of the IETF. OSPF is a link state protocol. Such protocols are also referred to in the literature as Shortest Path First-based or distributed-database protocols. As such, it sends routing information to all other routers within the same hierarchical area. Each OSPF router maintains an identical database describing the autonomous system's topology. From this database, a routing table is calculated by constructing a shortest-path tree using a dynamic (distributed) shortest path algorithm. OSPF is intended to recalculate routes quickly in the face of topological changes, utilizing a minimum of routing protocol traffic. As every router has the same database, the problem with loops that exist in DV routing protocols like RIP does not exist. OSPF provides support for equal-cost multi-path. An area routing capability is provided, enabling an additional level of routing protection and a reduction in routing protocol traffic. The first link-state routing protocol was developed for use in the ARPAnet packet switching network. This protocol formed the starting point for all other link-state protocols. The homogeneous ARPAnet environment, i.e. Single-vendor packet switches connected by synchronous serial lines simplified the design and implementation of the original protocol. The OSPF Working Group of the IETF has extended the original concepts in developing the OSPF protocol. The algorithms that they have developed have been tailored for efficient operation in TCP/IP internets.

> All routers run the exact same algorithm, in parallel. From the link-state database, each router constructs a tree of shortest paths with itself as root. This shortest-path tree gives the route to each destination in the Autonomous System. Externally derived routing information appears on the tree as leaves. When several equal-cost routes to a destination exist, traffic is distributed equally among them. The cost of a route is described by a single dimensionless metric. [10]

Information is promulgated around the network (AS) by a process of flooding. The main parameter that is required for this protocol is the "distance" metric or cost. The cost of a link must be between 1 and 65,535. The cost of a path is the sum of all covered links, but there is no explicit defined limit on path cost. OSPF selects the path with the lowest cost. The network administrator has complete control over the units and semantics of interface cost. For example, if each interface is assigned a cost of 1, OSPF find minimum-hop paths (just like RIP). If, instead, each interface is assigned a cost of the length in kilometers of the underlying physical circuit, OSPF will calculate paths having minimum static delay [15]. Reference [10] describes the basic shortest path procedure that can be shown to be a modification of Dijkstra's shortest path algorithm. Unlike RIP, OSPF uses IP directly, OSPF packets being identified by a special value in the IP datagram protocol field. OSPF is the recommended IPG for IPv6. In CISCO Routers the cost of a link is calculated with the formulae $\frac{100{,}000{,}000\text{bps}}{\text{bandwidth in bps}}$. This simple metric does not have the extensive set of weights used by EIGRP, but its possible to expand the metric to satisfy different QoS parameters, because no explicit formulae is defined in the standard. There's already an existing draft in the IETF which tries to expand the OSPF protocol. Aim of these draft is to introduce QoS with the minimum impact to the existing routing infrastructure. Thus, the authors enhanced the routing and topology data base with a QoS routing table. An interesting succession in the draft is the translation from delay requirements to bandwidth requirements. The authors use the expression $b(h) = \frac{a(h)}{D-h \cdot d}$ where $a(h) = \sigma + h \cdot c$, σ is the burst size, h is the number of hops on the path, c is the maximum packet size and D is the delay requirement. With the assumption that the propagation delays d_l of all links can be reasonably upperbounded by a single value d. After this translation the QoS metric depends only on the bandwidth parameter. For a further discussion see [27].

2.6 Exterior Gateway Protocol (EGP)

An exterior gateway protocol is designed to route between routing domains or better autonomous systems. Neither a distance-vector protocol, such as used by RIP, nor a link-state protocol, such as used by OSPF, is effective for an exterior routing protocol.

2.6.1 Path-vector Routing

The path-vector approach differs from a DV-algorithm in two respects:

- The path-vector approach does not include a distance or cost estimate.
- Each block of routing information lists all of the AS visited in order to reach the destination network by this route. The path information enables a router to perform policy routing. For example, information that is confidential may be limited to certain kinds of AS's.

2.6.2 Border Gateway Protocol (BGP)

BGP is an improvement over EGP. BGP was designed to allow special routers, called gateways, in different AS's to cooperate in the exchange of routing information [11]. BGP-4 is the current exterior routing protocol used for the global Internet. BGP uses TCP on port 179 as its transport protocol. On connection start, BGP peers exchange complete copies of their routing tables, which can be quite large. After the starting phase were the peers exchange the whole routing database, only changes are transmitted.

Three functional procedures are involved in BGP:

- Neighbour acquisition
  occurs when two neighbouring routers in different AS agree to exchange routing information.

- Neighbour reachability
  is used to maintain the relationship. Each partner needs to be assured that the other partner still exists and is still engaged in the neighbour relationship.

- Network reachability
  Each router maintains a database called RIB-In (Routing Information Base-Inbound) of the subnetworks that it can reach and the preferred route for reaching that subnetwork. BGP stores alternative routes in the database, so the size of the RIB-In can be many times the size of the router's routing table. Whenever a change is made to this database, the router issues an update message. By the broadcasting of these update message, all of the BGP routers can build up and maintain routing information.

To avoid loops, BGP routers employ a simple mechanism: When advertising a new destination, the whole path to the destination is included. Since the purpose of the BGP protocol is to exchange routing information between AS's, the complete path consists of a sequence of traversed AS's. A router avoid loops by never accepting an advertised destination if the associated path already includes the router's own AS number.

To fulfill QoS constrains between AS's one should take into consideration, that for reasons of both scalability and security, AS's will not reveal details of their internal structure to nodes outside. Instead, these AS's will only advertise a summary of their internal structure. The need for aggregation is obvious, even if one considers only scalability issues, since the computation and communication complexity of routing protocols grows at least linearly in the number of links in the network representation. There exist many aggregation schemes for EGP-Protocols. One possible choice has been recently adopted by the Private Network-to-Network Interface (PNNI) group of the ATM Forum. For some further work see [25] or [26].

	RIP	**IGRP**	**EIGRP**	**OSPF**	**BGP**
Static vs. Dynamic	D	D	D	D	D
Distributed vs. Centralized	D	D	D	D	D
Algorithm	DV	DV	DUAL	SPF	DV/PV
Single vs. Multi-Path	M	M	M	M	S
Flat vs. Hierarchical	F	F	F	H	F
Source vs. Router Intelligent	R	R	R	R	R
Type	IGP	IGP	IGP	IGP	EGP
Encapsulation	UDP(520)	IP(88)	IP(88)	IP(89)	TCP(179)
Security	no	no	yes	yes	yes
Neighbour discovery	no, only routing update	no	multicast Hello packets	multicast Hello packets	configured
Link State vs. Distance Factor	DF	DF	DF	LS	Neither (Reachability)
Variable netmask	no	no	yes	yes	yes
Metric Factors considered	Hop-Count	cost function (speed)	cost function (speed)	arbitrary[4]	arbitrary
Metric Ranges	1-16			1-65536	not specified

Table 2: Survey of the IP Routing Protocols

Algorithms

DV	=	Distance Vector (Bellman-Ford, Backward Search Algorithm)
DUAL	=	DV with diffusing update algorithm (Garcia-Luna-Aceves et al)
PV	=	"Path Vector"
SPF	=	Shortest-path-first (Dijkstra, Forward Search Algorithm)

3 Quality of Service and Constraint Based Routing (CBR)

Most of the multicast applications are real-time applications like video and audio. But the unicast and multicast routing algorithms do not support any kind of guarantees for bandwidth, delay

[4]In CISCO routers: $\frac{100{,}000{,}000\text{bps}}{\text{bandwidth in bps}}$

e.g. T1: $\frac{100{,}000{,}000\text{bps}}{1{,}544{,}000\text{bps}} = 64$

Ethernet: $\frac{100{,}000{,}000\text{bps}}{10\text{Mbps}} = 10$

or jitter. The IETF has proposed many service models and mechanisms to meet the demand for QoS. Notably among them are the Integrated Services/RSVP model [20, 21], MPLS and the Differentiated Services (DS) model [22, 23]. Thus it is obvious what should be the next step in the development of routing algorithms: QoS or Constraint Based Routing. In this section, we present some ideas and list some important constraints for QoS routing in the internet. For a more detailed framework see [19].

3.1 Performance Requirements in IP Networks

To compare different QoS Routing Protocols defined performance parameters are needed.

Basic performance parameters for any packet switching network include:

- Data throughput (try to maximize)
- Data transfer delay (try to minimize)
- Jitter

Quality of Service performance parameters for such networks often involve:

- Probability of call set-up failure
- Call set up delay
- Probability of failure during a call
- Probability of misdelivered calls/packets
- Call clear-down delay
- Probability of detectable and correctable errors
- Probability of detectable but not correctable errors
- Probability of undetected errors

3.2 Requirements for QoS Routing in IP Networks

3.2.1 Type of Service(ToS)-Routing

Before thinking about an optimal QoS Routing strategy, a certain amount of preliminary work needs to be done. What traffic classes exist in the network and what are their requirements? Should there be only two classes, e.g. stream-based and elastic traffic like in [13, 14], or three traffic classes (premium, assured, best effort) as discussed in the IETF DiffServ working group, or should there even be more? What are the appropriate parameters to describe the demands of the different traffic classes? One thing is clear, that different traffic classes need different routing strategies, thus a ToS Routing strategy is needed.

3.2.2 Marking of traffic streams and reservation

If the traffic classes are defined, the problem how to mark them arise. There is a flow-label field in IPv6 and in IPv4 it's possible to reuse the ToS-field, but the exact definition of the label is under discussion. Some traffic classes might need a reservation of resources. Is RSVP enough for this purpose or should be a extended or even new reservation mechanism deployed?

3.2.3 Metric

We have introduced EIGRP as a possibility for a complex metric. As already demonstrated in [28], most mixed metrics are NP-complete. Thus, a trade off between such a complex metric as in EIGRP and a simpler metric such as QOSPF may be needed. Further studies must be done to lower the computational complexity and describe the link parameters in an adequate way. In some situations it would seem that the load should be part of the metric used.

3.2.4 Adaptive-Routing-Algorithms

"Adaptive" routing has a long history in circuit-switched networks. It is instructive to review the adaptive routing methodologies like DAR [4] or RTNR , both to understand the problems encountered and possible solutions. There is for example much to be learned about alternate routing and its control and about dynamic resource sharing among different classes of traffic [3, 5] . Work in the area of ATM network routing shows that you must be very careful and not simply apply some of the results to a general topology network with heterogeneous multirate traffic.

3.2.5 Aggregation

For reasons of scalability and security, ISP will not reveal details of their internal network structure to outside nodes. Thus, these domains will only advertise a aggregated view of their internal structure. But what is the best aggregation scheme? Which parameter should be distributed? How precise are transferred informations about the routing parameter, e.g. load, and how often should they transferred?

3.2.6 Overhead of QoS Routing Protocols

4 Conclusions and Outlook

We have seen the evolution of routing schemes from simple static types to advanced dynamic routing schemes with a wide range of user-specifiable parameters. In the midst of such a bewildering array of new protocols and routing strategies in the Internet and high-speed networks such as ATM or SDH, it is important to be able to:

- analyze their performance accurately,
- predict where traffic streams will be directed,
- make intelligent choices for parameters on the basis of the traffic types involved,
- plan telecommunication networks adequately for the future.

From the work described in this paper, we can identify a number of substantial projects of importance and relevance to users of IP packet traffic:

1. The development of analytical tools to determine the performance of network routers employing the various gateway protocols described in this paper.
2. The development of a tool to enable the optimization of the various parameters K_1, K_2, ... K_5 used in the EIGRP protocol. Since, at present, the protocol offers many degrees of freedom but there is very little advice on the optimal selection of these parameters for the different classes of IP traffic (e.g. in the streaming or elastic categories).
3. The development of a new or modified version of a gateway protocol that addresses the problem of optimizing the routing for the various classes of traffic currently envisaged for the network.
4. Development of a complete Intranet Planning System with appropriate network design tools using advanced computer software incorporating tailored algorithms and supporting simulation programs.

References

[1] G.R. Ash, R.H. Cardwell, R.P. Murray, "Design and Optimisation of Networks with Dynamic Routing", Bell System Technical Journal, Vol. 60, No.8, Oct 1981

[2] G.R. Ash, A.H. Kafker, K.R. Krishnan, "Servicing and Real Time Control of Networks with Dynamic Routing", Bell System Technical Journal, Vol. 60, No.8, Oct 1981

[3] K.R. Krishnan and T.J. Ott, "State-Dependent Routing for Telephone Traffic: Theory and Results", Proc of the 25th Conf. on Decision and Control, Athens, Greece, Dec 1988

[4] R.J. Gibbens, F.P. Kelly, P.B. Key, "Dynamic Alternative Routing - Modelling and Behaviour", Proc. 12th International Teletraffic Congress, Torino, Italy, 1988

[5] K. Kawashima and A. Inoue, "State and Time Dependent Routing in the NTT Network", IEEE Communications Magazine, July 1995, pp40-47

[6] R.J. Harris, "A Course in Teletraffic Engineering", Telecom Australia, 1978 (Revised version under development)

[7] C. Hedrick, "Routing Information Protocol" RFC 1058, June, 1988

[8] G. Malkin, "RIP Version 2" RFC 2453, Nov, 1998

[9] J. Moy "OSPF Version 2" RFC 1247, July, 1991

[10] J. Moy, "OSPF Version 2" RFC 2328, April, 1998

[11] Y. Rekhter, T. Li "A Border Gateway Protocol 4 (BGP-4)", RFC 1771, March, 1995

[12] J.J. Garcia-Luna-Aceves, "Loop-Free Routing Using Diffusing Computations", IEEE/ACM Transactions on Networking, vol. 1, no. 1, 130-141, Feb, 1993

[13] J. Roberts, "Realizing Quality of Service Guarantees in Multiservice Networks", COST 257 Technical Document, 257TD(98)23, 1998

[14] J. Roberts, "Quality of Service Guarantees and Charging in Multi-service Networks", IEICE Transactions on Communications, 1998

[15] J.T. Moy, "OSPF: anatomy of an Internet routing protocol", Addison Wesley, 1998

[16] C. Huitema, "Routing in the Internet", Prentice Hall, 1995

[17] W. Stallings, "High-Speed Networks: TCP/IP and ATM design principles", Prentice-Hall, 1998

[18] S. Floyd, V. Jacobson "The Synchronization of Periodic Routing Messages", IEEE/ACM Transactions on Networking, vol.2 no. 2, 122-136, April, 1994

[19] E. Crawley, R. Nair, B. Rajagopalan, H. Sandick "A Framework for QoS-based Routing in the Internet" RFC 2386, Aug, 1998

[20] R. Braden, D. Clark, S. Shenker, "Integrated Services in the Internet Architecture: an Overview", RFC 1633, June, 1994

[21] R. Braden, L. Zhang, S. Berson, S. Herzog, S. Jamin, "Resource ReSerVation Protocol (RSVP) - Version 1 Functional Specification", RFC 2205, Sept, 1997

[22] K. Nichols, S. Blake, F. Baker, D. Black, “Definition of the Differentiated Service Field (DS Field) in the IPv4 and IPv6 Headers”, RFC 2474, Dec, 1998

[23] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, W. Weiss, “An Architecture for Differentieated Services”, RFC 2475, Dec, 1998

[24] C. Hedrick, “An Introduction to IGRP”, Rutgers - The State University of New Jersy Technical Publication, Laboratory for Computer Science Research, Aug, 1991

[25] Baruch Awerbuch, Yi Du, B. Kahn, Y. Shavitt, “Routing Through Networks with Hierarchical Topology Aggregation”, IEEE ISCC'98, Athens, Greece, 1998

[26] D. Lorenz, A. Orda, “QoS Routing in Networks with Uncertain Parameters” Proc. of INFOCOM'98, 1998

[27] G. Apostolopoulos et al, “QoS Routing Mechanisms and OSPF Extensions”, draft-guerin-qos-routing-ospf-04, Dec, 1998

[28] Z. Wang, J. Crowcroft “Quality of Service Routing for Supporting Multimedia Applications” IEEE JSAC, vol. 14, no. 7, Sep, 1996