Network Working Group   E. Crawley
Request for Comments: 2386   Argon Networks
Category: Informational   R. Nair
                                                         Arrowpoint
                                                   B. Rajagopalan
                                                            NEC USA
                                                         H. Sandick
                                                       Bay Networks
                                                        August 1998

             A Framework for QoS-based Routing in the Internet

Status of this Memo

Copyright Notice

ABSTRACT

   QoS-based routing has been recognized as a missing piece in the
   evolution of QoS-based service offerings in the Internet. This
   document describes some of the QoS-based routing issues and
   requirements, and proposes a framework for QoS-based routing in the
   Internet. This framework is based on extending the current Internet
   routing model of intra and interdomain routing to support QoS.

1. SCOPE OF DOCUMENT & PHILOSOPHY

   This document proposes a framework for QoS-based routing, with the
   objective of fostering the development of an Internet-wide solution
   while encouraging innovations in solving the many problems that
   arise. QoS-based routing has many complex facets and it is
   recommended that the following two-pronged approach be employed
   towards its development:

    1. Encourage the growth and evolution of novel intradomain QoS-based
       routing architectures. This is to allow the development of
       independent, innovative solutions that address the many QoS-based
       routing issues. Such solutions may be deployed in autonomous
       systems (ASs), large and small, based on their specific needs.

2. Encourage simple, consistent and stable interactions between ASs
   implementing routing solutions developed as above.

This approach follows the traditional separation between intra and
interdomain routing. It allows solutions like QOSPF [GKOP98, ZSSC97],
Integrated PNNI [IPNNI] or other schemes to be deployed for
intradomain routing without any restriction, other than their ability
to interact with a common, and perhaps simple, interdomain routing
protocol. The need to develop a single, all encompassing solution to
the complex problem of QoS-based routing is therefore obviated. As a
practical matter, there are many different views on how QoS-based
routing should be done. Much overall progress can be made if an
opportunity exists for various ideas to be developed and deployed
concurrently, while some consensus on the interdomain routing
architecture is being developed. Finally, this routing model is
perhaps the most practical from an evolution point of view. It is
superfluous to say that the eventual success of a QoS-based Internet
routing architecture would depend on the ease of evolution.

The aim of this document is to describe the QoS-based routing issues,
identify basic requirements on intra and interdomain routing, and
describe an extension of the current interdomain routing model to
support QoS. It is not an objective of this document to specify the
details of intradomain QoS-based routing architectures. This is left
up to the various intradomain routing efforts that might follow. Nor
is it an objective to specify the details of the interface between
reservation protocols such as RSVP and QoS-based routing. The
specific interface functionality needed, however, would be clear from
the intra and interdomain routing solutions devised. In the
intradomain area, the goal is to develop the basic routing
requirements while allowing maximum freedom for the development of
solutions. In the interdomain area, the objectives are to identify
the QoS-based routing functions, and facilitate the development or
enhancement of a routing protocol that allows relatively simple
interaction between domains.

In the next section, a glossary of relevant terminology is given. In
Section 3, the objectives of QoS-based routing are described and the
issues that must be dealt with by QoS-based Internet routing efforts
are outlined. In Section 4, some requirements on intradomain routing
are defined. These requirements are purposely broad, putting few
constraints on solution approaches. The interdomain routing model and
issues are described in Section 5 and QoS-based multicast routing is
discussed in Section 6. The interaction between QoS-based routing
and resource reservation protocols is briefly considered in Section
7. Security considerations are listed in Section 8 and related work
is described in Section 9. Finally, summary and conclusions are
presented in Section 10.

2. GLOSSARY

   The following glossary lists the terminology used in this document
   and an explanation of what is meant. Some of these terms may have
   different connotations, but when used in this document, their meaning
   is as given.

   Alternate Path Routing : A routing technique where multiple paths,
   rather than just the shortest path, between a source and a
   destination are utilized to route traffic. One of the objectives of
   alternate path routing is to distribute load among multiple paths in
   the network.

   Autonomous System (AS): A routing domain which has a common
   administrative authority and consistent internal routing policy. An
   AS may employ multiple intradomain routing protocols internally and
   interfaces to other ASs via a common interdomain routing protocol.

   Source: A host or router that can be identified by a unique unicast
   IP address.

   Unicast destination: A host or router that can be identified by a
   unique unicast IP address.

   Multicast destination: A multicast IP address indicating all hosts
   and routers that are members of the corresponding group.

   IP flow (or simply "flow"): An IP packet stream from a source to a
   destination (unicast or multicast) with an associated Quality of
   Service (QoS) (see below) and higher level demultiplexing
   information. The associated QoS could be "best-effort".

   Quality-of-Service (QoS): A set of service requirements to be met by
   the network while transporting a flow.

   Service class: The definitions of the semantics and parameters of a
   specific type of QoS.

   Integrated services: The Integrated Services model for the Internet
   defined in RFC 1633 allows for integration of QoS services with the
   best effort services of the Internet. The Integrated Services
   (IntServ) working group in the IETF has defined two service classes,
   Controlled Load Service [W97] and Guaranteed Service [SPG97].

   RSVP: The ReSerVation Protocol [BZBH97]. A QoS signaling protocol
   for the Internet.

   Path: A unicast or multicast path.

Unicast path: A sequence of links from an IP source to a unicast IP destination, determined by the routing scheme for forwarding packets.

Multicast path (or Multicast Tree): A subtree of the network topology in which all the leaves and zero or more interior nodes are members of the same multicast group. A multicast path may be per-source, in which case the subtree is rooted at the source.

Flow set-up: The act of establishing state in routers along a path to satisfy the QoS requirement of a flow.

Crankback: A technique where a flow setup is recursively backtracked along the partial flow path up to the first node that can determine an alternative path to the destination.

QoS-based routing: A routing mechanism under which paths for flows are determined based on some knowledge of resource availability in the network as well as the QoS requirement of flows.

Route pinning: A mechanism to keep a flow path fixed for a duration of time.

Flow Admission Control (FAC): A process by which it is determined whether a link or a node has sufficient resources to satisfy the QoS required for a flow. FAC is typically applied by each node in the path of a flow during flow set-up to check local resource availability.

Higher-level admission control: A process by which it is determined whether or not a flow set-up should proceed, based on estimates and policy requirements of the overall resource usage by the flow. Higher-level admission control may result in the failure of a flow set-up even when FAC at each node along the flow path indicates resource availability.

3. QOS-BASED ROUTING: BACKGROUND AND ISSUES

3.1 Best-Effort and QoS-Based Routing

   Routing deployed in today's Internet is focused on connectivity and
   typically supports only one type of datagram service called "best
   effort" [WC96]. Current Internet routing protocols, e.g. OSPF, RIP,
   use "shortest path routing", i.e. routing that is optimized for a
   single arbitrary metric, administrative weight or hop count. These
   routing protocols are also "opportunistic," using the current
   shortest path or route to a destination. Alternate paths with
   acceptable but non-optimal cost can not be used to route traffic
   (shortest path routing protocols do allow a router to alternate among

several equal cost paths to a destination).

QoS-based routing must extend the current routing paradigm in three basic ways. First, to support traffic using integrated-services class of services, multiple paths between node pairs will have to be calculated. Some of these new classes of service will require the distribution of additional routing metrics, e.g. delay, and available bandwidth. If any of these metrics change frequently, routing updates can become more frequent thereby consuming network bandwidth and router CPU cycles.

Second, today's opportunistic routing will shift traffic from one path to another as soon as a "better" path is found. The traffic will be shifted even if the existing path can meet the service requirements of the existing traffic. If routing calculation is tied to frequently changing consumable resources (e.g. available bandwidth) this change will happen more often and can introduce routing oscillations as traffic shifts back and forth between alternate paths. Furthermore, frequently changing routes can increase the variation in the delay and jitter experienced by the end users.

Third, as mentioned earlier, today's optimal path routing algorithms do not support alternate routing.   If the best existing path cannot admit a new flow, the associated traffic cannot be forwarded even if an adequate alternate path exists.

3.2 QoS-Based Routing and Resource Reservation

It is important to understand the difference between QoS-based routing and resource reservation. While resource reservation protocols such as RSVP [BZBH97] provide a method for requesting and reserving network resources, they do not provide a mechanism for determining a network path that has adequate resources to accommodate the requested QoS. Conversely, QoS-based routing allows the determination of a path that has a good chance of accommodating the requested QoS, but it does not include a mechanism to reserve the required resources.

Consequently, QoS-based routing is usually used in conjunction with some form of resource reservation or resource allocation mechanism. Simple forms of QoS-based routing have been used in the past for Type of Service (TOS) routing [M98]. In the case of OSPF, a different shortest-path tree can be computed for each of the 8 TOS values in the IP header [ISI81]. Such mechanisms can be used to select specially provisioned paths but do not completely assure that resources are not overbooked along the path. As long as strict resource management and control are not needed, mechanisms such as TOS-based routing are useful for separating whole classes of traffic

over multiple routes. Such mechanisms might work well with the
emerging Differential Services efforts [BBCD98].

Combining a resource reservation protocol with QoS-based routing
allows fine control over the route and resources at the cost of
additional state and setup time. For example, a protocol such as RSVP
may be used to trigger QoS-based routing calculations to meet the
needs of a specific flow.

3.3 QoS-Based Routing: Objectives

Under QoS-based routing, paths for flows would be determined based
on some knowledge of resource availability in the network, as well as
the QoS requirement of flows. The main objectives of QoS-based
routing are:

1. Dynamic determination of feasible paths: QoS-based routing can
    determine a path, from among possibly many choices, that has a
    good chance of accommodating the QoS of the given flow. Feasible
    path selection may be subject to policy constraints, such as path
    cost, provider selection, etc.

2. Optimization of resource usage: A network state-dependent QoS-
    based routing scheme can aid in the efficient utilization of
    network resources by improving the total network throughput. Such
    a routing scheme can be the basis for efficient network
    engineering.

3. Graceful performance degradation: State-dependent routing can
    compensate for transient inadequacies in network engineering
    (e.g., during focused overload conditions), giving better
    throughput and a more graceful performance degradation as
    compared to a state-insensitive routing scheme [A84].

QoS-based routing in the Internet, however, raises many issues:

- How do routers determine the QoS capability of each outgoing link
   and reserve link resources? Note that some of these links may be
   virtual, over ATM networks and others may be broadcast multi-
   access links.

- What is the granularity of routing decision (i.e., destination-
   based, source and destination-based, or flow-based)?

- What routing metrics are used and how are QoS-accommodating paths
   computed for unicast flows?

- How are QoS-accommodating paths computed for multicast flows with different reservation styles and receiver heterogeneity?

- What are the performance objectives while computing QoS-based paths?

- What are the administrative control issues?

- What factors affect the routing overheads?, and

- How is scalability achieved?

Some of these issues are discussed briefly next. Interdomain routing is discussed in Section 5.

3.4 QoS Determination and Resource Reservation

To determine whether the QoS requirements of a flow can be accommodated on a link, a router must be able to determine the QoS available on the link. It is still an open issue as to how the QoS availability is determined for broadcast multiple access links (e.g., Ethernet). A related problem is the reservation of resources over such links. Solutions to these problems are just emerging [GPSS98].

Similar problems arise when a router is connected to a large non-broadcast multiple access network, such as ATM. In this case, if the destination of a flow is outside the ATM network, the router may have multiple egress choices. Furthermore, the QoS availability on the ATM paths to each egress point may be different. The issues then are,

    o   how does a router determine all the egress choices across the
         ATM network?
    o   how does it determine what QoS is available over the path to
         each egress point?, and
    o   what QoS value does the router advertise for the ATM link.

Typically, IP routing over ATM (e.g., NHRP) allows the selection of a single egress point in the ATM network, and the procedure does not incorporate any knowledge of the QoS required over the path. An approach like I-PNNI [IPNNI] would be helpful here, although it introduces some complexity.

An additional problem with resource reservation is how to determine what resources have already been allocated to a multicast flow. The availability of this information during path computation improves the chances of finding a path to add a new receiver to a multicast flow. QOSPF [ZSSC97] handles this problem by letting routers broadcast reserved resource information to other routers in their area.

Alternate path routing [ZES97] deals with this issue by using probe
messages to find a path with sufficient resources. Path QoS
Computation (PQC) method, proposed in [GOA97], propagates bandwidth
allocation information in RSVP PATH messages. A router receiving the
PATH message gets an indication of the resource allocation only on
those links in the path to itself from the source. Allocation for
the same flow on other remote branches of the multicast tree is not
available. Thus, the PQC method may not be sufficient to find
feasible QoS-accommodating paths to all receivers.

3.5 Granularity of Routing Decision

Routing in the Internet is currently based only on the destination
address of a packet. Many multicast routing protocols require
routing based on the source AND destination of a packet. The
Integrated Services architecture and RSVP allow QoS determination for
an individual flow between a source and a destination. This set of
routing granularities presents a problem for QoS routing solutions.

If routing based only on destination address is considered, then an
intermediate router will route all flows between different sources
and a given destination along the same path. This is acceptable if
the path has adequate capacity but a problem arises if there are
multiple flows to a destination that exceed the capacity of the link.

One version of QOSPF [ZSSC97] determines QoS routes based on source
and destination address. This implies that all traffic between a
given source and destination, regardless of the flow, will travel
down the same route. Again, the route must have capacity for all the
QoS traffic for the source/destination pair. The amount of routing
state also increases since the routing tables must include
source/destination pairs instead of just the destination.

The best granularity is found when routing is based on individual
flows but this incurs a tremendous cost in terms of the routing
state. Each QoS flow can be routed separately between any source and
destination. PQC [GOA97] and alternate path routing [ZES97], are
examples of solutions which operate at the flow level.

Both source/destination and flow-based routing may be susceptible to
packet looping under hop-by-hop forwarding. Suppose a node along a
flow or source/destination-based path loses the state information for
the flow. Also suppose that the flow-based route is different from
the regular destination-based route. The potential then exists for a
routing loop to form when the node forwards a packet belonging to the
flow using its destination-based routing table to a node that occurs

earlier on the flow-based path. This is because the latter node may
use its flow-based routing table to forward the packet again to the
former and this can go on indefinitely.

3.6   Metrics and Path Computation

3.6.1 Metric Selection and Representation

There are some considerations in defining suitable link and node
metrics [WC96]. First, the metrics must represent the basic network
properties of interest. Such metrics include residual bandwidth,
delay and jitter. Since the flow QoS requirements have to be mapped
onto path metrics, the metrics define the types of QoS guarantees the
network can support. Alternatively, QoS-based routing cannot support
QoS requirements that cannot be meaningfully mapped onto a reasonable
combination of path metrics. Second, path computation based on a
metric or a combination of metrics must not be too complex as to
render them impractical. In this regard, it is worthwhile to note
that path computation based on certain combinations of metrics (e.g.,
delay and jitter) is theoretically hard. Thus, the allowable
combinations of metrics must be determined while taking into account
the complexity of computing paths based on these metrics and the QoS
needs of flows. A common strategy to allow flexible combinations of
metrics while at the same time reduce the path computation complexity
is to utilize "sequential filtering". Under this approach, a
combination of metrics is ordered in some fashion, reflecting the
importance of different metrics (e.g., cost followed by delay, etc.).
Paths based on the primary metric are computed first (using a simple
algorithm, e.g., shortest path) and a subset of them are eliminated
based on the secondary metric and so forth until a single path is
found. This is an approximation technique and it trades off global
optimality for path computation simplicity (The filtering technique
may be simpler, depending on the set of metrics used. For example,
with bandwidth and cost as metrics, it is possible to first eliminate
the set of links that do not have the requested bandwidth and then
compute the least cost path using the remaining links.)

Now, once suitable link and node metrics are defined, a uniform
representation of them is required across independent domains -
employing possibly different routing schemes - in order to derive
path metrics consistently (path metrics are obtained by the
composition of link and node metrics). Encoding of the maximum,
minimum, range, and granularity of the metrics are needed. Also, the
definitions of comparison and accumulation operators are required. In
addition, suitable triggers must be defined for indicating a
significant change from a minor change. The former will cause a
routing update to be generated. The stability of the QoS routes would

depend on the ability to control the generation of updates. With
interdomain routing, it is essential to obtain a fairly stable view
of the interconnection among the ASs.

3.6.2 Metric Hierarchy

A hierarchy can be defined among various classes of service based on
the degree to which traffic from one class can potentially degrade
service of traffic from lower classes that traverse the same link. In
this hierarchy, guaranteed constant bit rate traffic is at the top
and "best-effort" datagram traffic at the bottom. Classes providing
service higher in the hierarchy impact classes providing service in
lower levels. The same situation is not true in the other direction.
For example, a datagram flow cannot affect a real-time service. Thus,
it may be necessary to distribute and update different metrics for
each type of service in the worst case. But, several advantages
result by identifying a single default metric. For example, one
could derive a single metric combining the availability of datagram
and real-time service over a common substrate.

3.6.3 Datagram Flows

A delay-sensitive metric is probably the most obvious type of metric
suitable for datagram flows. However, it requires careful analysis to
avoid instabilities and to reduce storage and bandwidth requirements.
For example, a recursive filtering technique based on a simple and
efficient weighted averaging algorithm [NC94] could be used. This
filter is used to stabilize the metric. While it is adequate for
smoothing most loading patterns, it will not distinguish between
patterns consisting of regular bursts of traffic and random loading.
Among other stabilizing tools, is a minimum time between updates that
can help filter out high-frequency oscillations.

3.6.4 Real-time Flows

In real-time quality-of-service, delay variation is generally more
critical than delay as long as the delay is not too high. Clearly,
voice-based applications cannot tolerate more than a certain level of
delay. The condition of varying delays may be expected to a greater
degree in a shared medium environment with datagrams, than in a
network implemented over a switched substrate. Routing a real-time
flow therefore reduces to an exercise in allocating the required
network resources while minimizing fragmentation of bandwidth. The
resulting situation is a bandwidth-limited minimum hop path from a
source to the destination. In other words, the router performs an
ordered search through paths of increasing hop count until it finds
one that meets all the bandwidth needs of the flow. To reduce
contention and the probability of false probes (due to inaccuracy in

route tables), the router could select a path randomly from a
"window" of paths which meet the needs of the flow and satisfy one of
three additional criteria: best-fit, first-fit or worst-fit. Note
that there is a similarity between the allocation of bandwidth and
the allocation of memory in a multiprocessing system. First-fit seems
to be appropriate for a system with a high real-time flow arrival
rates; and worst-fit is ideal for real-time flows with high holding
times. This rather nonintuitive result was shown in [NC94].

3.6.5 Path Properties

Path computation by itself is merely a search technique, e.g.,
Shortest Path First (SPF) is a search technique based on dynamic
programming. The usefulness of the paths computed depends to a large
extent on the metrics used in evaluating the cost of a path with
respect to a flow.

Each link considered by the path computation engine must be evaluated
against the requirements of the flow, i.e., the cost of providing the
services required by the flow must be estimated with respect to the
capabilities of the link. This requires a uniform method of combining
features such as delay, bandwidth, priority and other service
features. Furthermore, the costs must reflect the lost opportunity
of using each link after routing the flow.

3.6.6 Performance Objectives

One common objective during path computation is to improve the total
network throughput. In this regard, merely routing a flow on any
path that accommodates its QoS requirement is not a good strategy. In
fact, this corresponds to uncontrolled alternate routing [SD95] and
may adversely impact performance at higher traffic loads. It is
therefore necessary to consider the total resource allocation for a
flow along a path, in relation to available resources, to determine
whether or not the flow should be routed on the path. Such a
mechanism is referred to in this document as "higher level admission
control". The goal of this is to ensure that the "cost" incurred by
the network in routing a flow with a given QoS is never more than the
revenue gained. The routing cost in this regard may be the lost
revenue in potentially blocking other flows that contend for the same
resources. The formulation of the higher level admission control
strategy, with suitable administrative hooks and with fairness to all
flows desiring entry to the network, is an issue. The fairness
problem arises because flows with smaller reservations tend to be
more successfully routed than flows with large reservations, for a
given engineered capacity. To guarantee a certain level of

acceptance rate for "larger" flows, without over-engineering the
network, requires a fair higher level admission control mechanism.
The application of higher level admission control to multicast
routing is discussed later.

3.7   Administrative Control

There are several administrative control issues. First, within an AS
employing state-dependent routing, administrative control of routing
behavior may be necessary. One example discussed earlier was higher
level admission control. Some others are described in this section.
Second, the control of interdomain routing based on policy is an
issue. The discussion of interdomain routing is defered to Section
5.

Two areas that need administrative control, in addition to
appropriate routing mechanisms, are handling flow priority with
preemption, and resource allocation for multiple service classes.

3.7.1 Flow Priorities and Preemption

If there are critical flows that must be accorded higher priority
than other types of flows, a mechanism must be implemented in the
network to recognize flow priorities. There are two aspects to
prioritizing flows. First, there must be a policy to decide how
different users are allowed to set priorities for flows they
originate. The network must be able to verify that a given flow is
allowed to claim a priority level signaled for it. Second, the
routing scheme must ensure that a path with the requested QoS will be
found for a flow with a probability that increases with the priority
of the flow. In other words, for a given network load, a high
priority flow should be more likely to get a certain QoS from the
network than a lower priority flow requesting the same QoS. Routing
procedures for flow prioritization can be complex. Identification
and evaluation of different procedures are areas that require
investigation.

3.7.2 Resource Control

If there are multiple service classes, it is necessary to engineer a
network to carry the forecasted traffic demands of each class. To do
this, router and link resources may be logically partitioned among
various service classes. It is desirable to have dynamic partitioning
whereby unused resources in various partitions are dynamically
shifted to other partitions on demand [ACFH92]. Dynamic sharing,
however, must be done in a controlled fashion in order to prevent
traffic under some service class from taking up more resources than

what was engineered for it for prolonged periods of time. The design
of such a resource sharing scheme, and its incorporation into the
QoS-based routing scheme are significant issues.

3.8   QoS-Based Routing for Multicast Flows

QoS-based multicast routing is an important problem, especially if
the notion of higher level admission control is included. The
dynamism in the receiver set allowed by IP multicast, and receiver
heterogeneity add to the problem. With straightforward implementation
of distributed heuristic algorithms for multicast path computation
[W88, C91], the difficulty is essentially one of scalability. To
accommodate QoS, multicast path computation at a router must have
knowledge of not only the id of subnets where group members are
present, but also the identity of branches in the existing tree. In
other words, routers must keep flow-specific state information. Also,
computing optimal shared trees based on the shared reservation style
[BZBH97], may require new algorithms. Multicast routing is discussed
in some detail in Section 6.

3.9   Routing Overheads

The overheads incurred by a routing scheme depend on the type of the
routing scheme, as well as the implementation. There are three types
of overheads to be considered: computation, storage and
communication. It is necessary to understand the implications of
choosing a routing mechanism in terms of these overheads.

For example, considering link state routing, the choice of the update
propagation mechanism is important since network state is dynamic and
changes relatively frequently. Specifically, a flooding mechanism
would result in many unnecessary message transmissions and
processing. Alternative techniques, such as tree-based forwarding
[R96], have to be considered. A related issue is the quantization of
state information to prevent frequent updating of dynamic state.
While coarse quantization reduces updating overheads, it may affect
the performance of the routing scheme. The tradeoff has to be
carefully evaluated. QoS-based routing incurs certain overheads
during flow establishment, for example, computing a source route.
Whether this overhead is disproportionate compared to the length of
the sessions is an issue. In general, techniques for the minimization
of routing-related overheads during flow establishment must be
investigated. Approaches that are useful include pre-computation of
routes, caching recently used routes, and TOS routing based on hints
in packets (e.g., the TOS field).

3.10   Scaling by Hierarchical Aggregation

   QoS-based routing should be scalable, and hierarchical aggregation is
   a common technique for scaling (e.g., [PNNI96]). But this introduces
   problems with regard to the accuracy of the aggregated state
   information [L95]. Also, the aggregation of paths under multiple
   constraints is difficult. One of the difficulties is the risk of
   accepting a flow based on inaccurate information, but not being able
   to support the QoS requirements of flow because the capabilities of
   the actual paths that are aggregated are not known during route
   computation. Performance impacts of aggregating path metric
   information must therefore be understood. A way to compensate for
   inaccuracies is to use crankback, i.e., dynamic search for alternate
   paths as a flow is being routed. But crankback increases the time to
   set up a flow, and may adversely affect the performance of the
   routing scheme under some circumstances. Thus, crankback must be used
   judiciously, if at all, along with a higher level admission control
   mechanism.

4. INTRADOMAIN ROUTING REQUIREMENTS

   At the intradomain level, the objective is to allow as much latitude
   as possible in addressing the QoS-based routing issues. Indeed, there
   are many ideas about how QoS-based routing services can be
   provisioned within ASs. These range from on-demand path computation
   based on current state information, to statically provisioned paths
   supporting a few service classes.

   Another aspect that might invite differing solutions is performance
   optimization. Based on the technique used for this, intradomain
   routing could be very sophisticated or rather simple. Finally, the
   service classes supported, as well as the specific QoS engineered for
   a service class, could differ from AS to AS. For instance, some ASs
   may not support guaranteed service, while others may. Also, some ASs
   supporting the service may be engineered for a better delay bound
   than others. Thus, it requires considerable thought to determine the
   high level requirements for intradomain routing that both supports
   the overall view of QoS-based routing in the Internet and allows
   maximum autonomy in developing solutions.

   Our view is that certain minimum requirements must be satisfied by
   intradomain routing in order to be qualified as "QoS-based" routing.
   These are:

   - The routing scheme must route a flow along a path that can
     accommodate its QoS requirements, or indicate that the flow cannot
     be admitted with the QoS currently being requested.

    - The routing scheme must indicate disruptions to the current route
      of a flow due to topological changes.

    - The routing scheme must accommodate best-effort flows without any
      resource reservation requirements. That is, present best effort
      applications and protocol stacks need not have to change to run in
      a domain employing QoS-based routing.

    - The routing scheme may optionally support QoS-based multicasting
      with receiver heterogeneity and shared reservation styles.

    In addition, the following capabilities are also recommended:

    - Capabilities to optimize resource usage.

    - Implementation of higher level admission control procedures to
      limit the overall resource utilization by individual flows.

    Further requirements along these lines may be specified. The
    requirements should capture the consensus view of QoS-based routing,
    but should not preclude particular approaches (e.g., TOS-based
    routing) from being implemented. Thus, the intradomain requirements
    are expected to be rather broad.

5. INTERDOMAIN ROUTING

    The fundamental requirement on interdomain QoS-based routing is
    scalability. This implies that interdomain routing cannot be based
    on highly dynamic network state information. Rather, such routing
    must be aided by sound network engineering and relatively sparse
    information exchange between independent routing domains. This
    approach has the advantage that it can be realized by straightforward
    extensions of the present Internet interdomain routing model. A
    number of issues, however, need to be addressed to achieve this, as
    discussed below.

5.1 Interdomain QoS-Based Routing Model

   The interdomain QoS-based routing model is depicted below:

```
            AS1                    AS2                  AS3
       _____        _____      _____
      |      |       |      |      |       |     |     |        |
      |      |   B------B   |      |   B----B    |     |        |
      |      |       |      |      |       |     |     |        |
       -----B-----        B-------------      --B---------
            \        /                       /
             \      /                       /
        _____B_____B_____          _____B_____
       |      |        |          |               |
       |      |        |          |               |
       |      |   B-------B       |               |
       |      |   |       |       |               |
       |      |   B-------B       |               |
       |      ---------------      ---------------
              AS4                        AS5
```
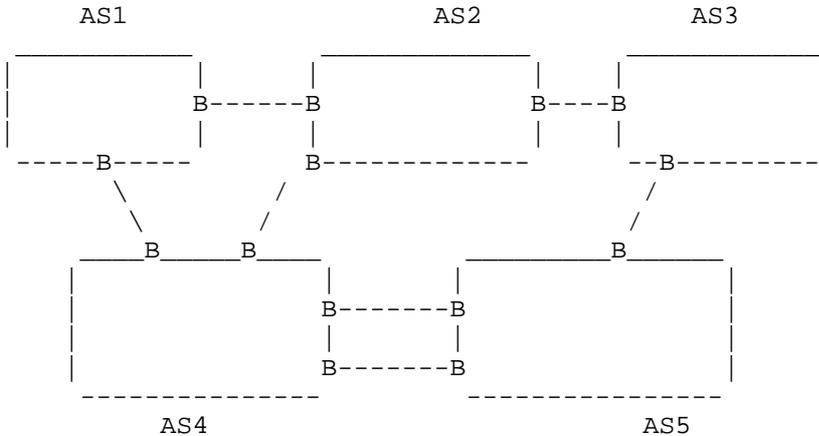
   Here, ASs exchange standardized routing information via border nodes
   B. Under this model, each AS can itself consist of a set of
   interconnected ASs, with standardized routing interaction. Thus, the
   interdomain routing model is hierarchical. Also, each lowest level
   AS employs an intradomain QoS-based routing scheme (proprietary or
   standardized by intradomain routing efforts such as QOSPF). Given
   this structure, some questions that arise are:

   - What information is exchanged between ASs?

   - What routing capabilities does the information exchange lead to?
     (E.g., source routing, on-demand path computation, etc.)

   - How is the external routing information represented within an AS?

   - How are interdomain paths computed?

   - What sort of policy controls may be exerted on interdomain path
     computation and flow routing?, and

   - How is interdomain QoS-based multicast routing accomplished?

   At a high level, the answers to these questions depend on the routing
   paradigm. Specifically, considering link state routing, the
   information exchanged between domains would consist of an abstract
   representation of the domains in the form of logical nodes and links,
   along with metrics that quantify their properties and resource
   availability. The hierarchical structure of the ASs may be handled

by a hierarchical link state representation, with appropriate metric
aggregation.

Link state routing may not necessarily be advantageous for
interdomain routing for the following reasons:

- One advantage of intradomain link state routing is that it would
  allow fairly detailed link state information be used to compute
  paths on demand for flows requiring QoS. The state and metric
  aggregation used in interdomain routing, on the other hand, erodes
  this property to a great degree.

- The usefulness of keeping track of the abstract topology and
  metrics of a remote domain, or the interconnection between remote
  domains is not obvious. This is especially the case when the remote
  topology and metric encoding are lossy.

- ASs may not want to advertise any details of their internal
  topology or resource availability.

- Scalability in interdomain routing can be achieved only if
  information exchange between domains is relatively infrequent.
  Thus, it seems practical to limit information flow between domains
  as much as possible.

Compact information flow allows the implementation QoS-enhanced
versions of existing interdomain protocols such as BGP-4. We look at
the interdomain routing issues in this context.

5.2 Interdomain Information Flow

The information flow between routing domains must enable certain
basic functions:

1. Determination of reachability to various destinations

2. Loop-free flow routes

3. Address aggregation whenever possible

4. Determination of the QoS that will be supported on the path to a
    destination. The QoS information should be relatively static,
    determined from the engineered topology and capacity of an AS
    rather than ephemeral fluctuations in traffic load through the
    AS. Ideally, the QoS supported in a transit AS should be allowed
    to vary significantly only under exceptional circumstances, such
    as failures or focused overload.

5. Determination, optionally, of multiple paths for a given
    destination, based on service classes.

6. Expression of routing policies, including monetary cost, as a
    function of flow parameters, usage and administrative factors.

Items 1-3 are already part of existing interdomain routing. Item 5 is
also a straightfoward extension of the current model. The main
problem areas are therefore items 4 and 6.

The QoS of an end-to-end path is obtained by composing the QoS
available in each transit AS. Thus, border routers must first
determine what the locally available QoS is in order to advertise
routes to both internal and external destinations. The determination
of local "AS metrics" (corresponding to link metrics in the
intradomain case) should not be subject to too much dynamism. Thus,
the issue is how to define such metrics and what triggers an
occasional change that results in re-advertisements of routes.

The approach suggested in this document is not to compute paths based
on residual or instantaneous values of AS metics (which can be
dynamic), but utilize only the QoS capabilities engineered for
aggregate transit flows. Such engineering may be based on the
knowledge of traffic to be expected from each neighboring ASs and the
corresponding QOS needs. This information may be obtained based on
contracts agreed upon prior to the provisioning of services. The AS
metric then corresponds to the QoS capabilities of the "virtual path"
engineered through the AS (for transit traffic) and a different
metric may be used for different neighbors. This is illustrated in
the following figure.

```
         AS1                      AS2                 AS3

       _____          _____      _____
      |           |        |      |      |    |           |
      |        B------B1    |     B2----B     |
      |        |   |        |      |    |      |           |
       -----B-----         B3------------      --B---------
           \       |          /
            \             /
             \          /
          ____B_____B_____
          |         |
          |         |
          |         |
          |         |
           --------------
              AS4
```
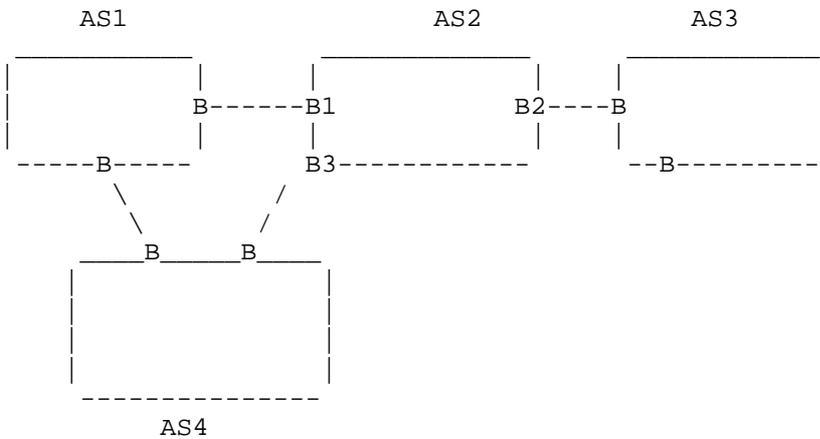
Here, B1 may utilize an AS metric specific for AS1 when computing
path metrics to be advertised to AS1. This metric is based on the
resources engineered in AS2 for transit traffic from AS1. Similarly,
B3 may utilize a different metric when computing path metrics to be
advertised to AS4. Now, it is assumed that as long as traffic flow
into AS2 from AS1 or AS4 does not exceed the engineered values, these
path metrics would hold. Excess traffic due to transient
fluctuations, however, may be handled as best effort or marked with a
discard bit.

Thus, this model is different from the intradomain model, where end
nodes pick a path dynamically based on the QoS needs of the flow to
be routed. Here, paths within ASs are engineered based on presumed,
measured or declared traffic and QoS requirements. Under this model,
an AS can contract for routes via multiple transit ASs with different
QoS requirements. For instance, AS4 above can use both AS1 and AS2 as
transits for same or different destinations. Also, a QoS contract
between one AS and another may generate another contract between the
second and a third AS and so forth.

An issue is what triggers the recomputation of path metrics within an
AS. Failures or other events that prevent engineered resource
allocation should certainly trigger recomputation. Recomputation
should not be triggered in response to arrival of flows within the
engineered limit.

5.3   Path Computation

Path computation for an external destination at a border node is
based on reachability, path metrics and local policies of selection.
If there are multiple selection criteria (e.g., delay, bandwidth,
cost, etc.), mutiple alternaives may have to be maintained as well as
propagated by border nodes. Selection of a path from among many
alternatives would depend on the QoS requests of flows, as well as
policies. Path computation may also utilze any heuristics for
optimizing resource usage.

5.4 Flow Aggregation

An important issue in interdomain routing is the amount of flow state
to be processed by transit ASs. Reducing the flow state by
aggregation techniques must therefore be seriously considered. Flow
aggregation means that transit traffic through an AS is classified
into a few aggregated streams rather than being routed at the
individual flow level. For example, an entry border router may
classify various transit flows entering an AS into a few coarse
categories, based on the egress node and QoS requirements of the
flows. Then, the aggregated stream for a given traffic class may be

routed as a single flow inside the AS to the exit border router. This
router may then present individual flows to different neighboring ASs
and the process repeats at each entry border router. Under this
scenario, it is essential that entry border routers keep track of the
resource requirements for each transit flow and apply admission
control to determine whether the aggregate requirement from any
neighbor exceeds the engineered limit. If so, some policy must be
invoked to deal with the excess traffic. Otherwise, it may be assumed
that aggregated flows are routed over paths that have adequate
resources to guarantee QoS for the member flows. Finally, it is
possible that entry border routers at a transit AS may prefer not to
aggregate flows if finer grain routing within the AS may be more
efficient (e.g., to aid load balancing within the AS).

5.5   Path Cost Determination

   It is hoped that the integrated services Internet architecture would
   allow providers to charge for IP flows based on their QoS
   requirements. A QoS-based routing architecture can aid in
   distributing information on expected costs of routing flows to
   various destinations via different domains. Clearly, from a
   provider's point of view, there is a cost incurred in guaranteeing
   QoS to flows. This cost could be a function of several parameters,
   some related to flow parameters, others based on policy. From a
   user's point of view, the consequence of requesting a particular QoS
   for a flow is the cost incurred, and hence the selection of providers
   may be based on cost. A routing scheme can aid a provider in
   distributing the costs in routing to various destinations, as a
   function of several parameters, to other providers or to end users.
   In the interdomain routing model described earlier, the costs to a
   destination will change as routing updates are passed through a
   transit domain. One of the goals of the routing scheme should be to
   maintain a uniform semantics for cost values (or functions) as they
   are handled by intermediate domains. As an example, consider the cost
   function generated by border node B1 in domain A and passed to node
   B2 in domain B below. The routing update may be injected into domain
   B by B2 and finally passed to B4 in domain C by router B3. Domain B
   may interpret the cost value received from domain A in any way it
   wants, for instance, adding a locally significant component to it.
   But when this cost value is passed to domain C, the meaning of it
   must be what domain A intended, plus the incremental cost of
   transiting domain B, but not what domain B uses internally.

```
    Domain A                    Domain B            Domain C
  _____      _____      _____
 |             |    |            |     |            |     |
 |             |    |            |     |            |     |
 |           B1|------|B2          |   B3|---|B4        |
 |             |    |            |     |            |     |
  ------------      ----------      ------------
```
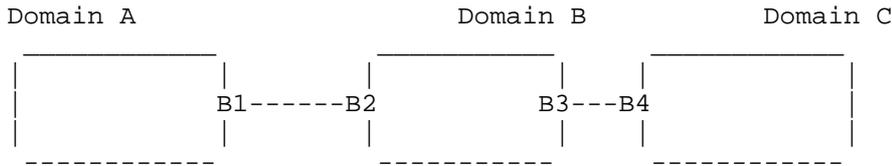
A problem with charging for a flow is the determination of the cost
when the QoS promised for the flow was not actually delivered.
Clearly, when a flow is routed via multiple domains, it must be
determined whether each domain delivers the QoS it declares possible
for traffic through it.

6. QOS-BASED MULTICAST ROUTING

   The goals of QoS-based multicast routing are as follows:

   - Scalability to large groups with dynamic membership

   - Robustness in the presence of topological changes

   - Support for receiver-initiated, heterogeneous reservations

   - Support for shared reservation styles, and

   - Support for "global" admission control, i.e., administrative
     control of resource consumption by the multicast flow.

   The RSVP multicast flow model is as follows. The sender of a
   multicast flow advertises the traffic characteristics periodically to
   the receivers. On receipt of an advertisement, a receiver may
   generate a message to reserve resources along the flow path from the
   sender. Receiver reservations may be heterogeneous. Other multicast
   models may be considered.

   The multicast routing scheme attempts to determine a path from the
   sender to each receiver that can accommodate the requested
   reservation. The routing scheme may attempt to maximize network
   resource utilization by minimizing the total bandwidth allocated to
   the multicast flow, or by optimizing some other measure.

6.1   Scalability, Robustness and Heterogeneity

   When addressing scalability, two aspects must be considered:

   1.   The overheads associated with receiver discovery. This overhead
        is incurred when determining the multicast tree for forwarding
        best-effort sender traffic characterization to receivers.

2. The overheads associated with QoS-based multicast path
    computation. This overhead is incurred when flow-specific
    state information has to be collected by a router to determine
    QoS-accommodating paths to a receiver.

Depending on the multicast routing scheme, one or both of these
aspects become important. For instance, under the present RSVP model,
reservations are established on the same path over which sender
traffic characterizations are sent, and hence there is no path
computation overhead. On the other hand, under the proposed QOSPF
model [ZSSC97] of multicast source routing, receiver discovery
overheads are incurred by MOSPF [M94] receiver location broadcasts,
and additional path computation overheads are incurred due to the
need to keep track of existing flow paths. Scaling of QoS-based
multicast depends on both these scaling issues. However, scalable
best-effort multicasting is really not in the domain of QoS-based
routing work (solutions for this are being devised by the IDMR WG
[BCF94, DEFV94]). QoS-based multicast routing may build on these
solutions to achieve overall scalability.

There are several options for QoS-based multicast routing. Multicast
source routing is one under which multicast trees are computed by the
first-hop router from the source, based on sender traffic
advertisements. The advantage of this is that it blends nicely with
the present RSVP signaling model. Also, this scheme works well when
receiver reservations are homogeneous and the same as the maximum
reservation derived from sender advertisement. The disadvantages of
this scheme are the extra effort needed to accommodate heterogeneous
reservations and the difficulties in optimizing resource allocation
based on shared reservations.

In these regards, a receiver-oriented multicast routing model seems
to have some advantage over multicast source routing. Under this
model:

1. Sender traffic advertisements are multicast over a best-effort
    tree which can be different from the QoS-accommodating tree for
    sender data.

2. Receiver discovery overheads are minimized by utilizing a
    scalable scheme (e.g., PIM, CBT), to multicast sender traffic
    characterization.

3. Each receiver-side router independently computes a QoS-
    accommodating path from the source, based on the receiver
    reservation. This path can be computed based on unicast routing
    information only, or with additional multicast flow-specific
    state information. In any case, multicast path computation is

broken up into multiple, concurrent nunicast path computations.

      4.   Routers processing unicast reserve messages from receivers
           aggregate resource reservations from multiple receivers.

Flow-specific state information may be limited in Step 3 to achieve
scalability [RN98]. In general, limiting flow-specific information in
making multicast routing decisions is important in any routing model.
The advantages of this model are the ease with which heterogeneous
reservations can be accommodated, and the ability to handle shared
reservations. The disadvantages are the incompatibility with the
present RSVP signaling model, and the need to rely on reverse paths
when link state routing is not used. Both multicast source routing
and the receiver-oriented routing model described above utilize per-
source trees to route multicast flows. Another possibility is the
utilization of shared, per-group trees for routing flows. The
computation and usage of such trees require further work.

Finally, scalability at the interdomain level may be achieved if
QoS-based multicast paths are computed independently in each domain.
This principle is illustrated by the QOSPF multicast source routing
scheme which allows independent path computation in different OSPF
areas. It is easy to incorporate this idea in the receiver-oriented
model also. An evaluation of multicast routing strategies must take
into account the relative advantages and disadvantages of various
approaches, in terms of scalability features and functionality
supported.

6.2   Multicast Admission Control

Higher level admission control, as defined for unicast, prevents
excessive resource consumption by flows when traffic load is high.
Such an admission control strategy must be applied to multicast flows
when the flow path computation is receiver-oriented or sender-
oriented. In essence, a router computing a path for a receiver must
determine whether the incremental resource allocation for the
receiver is excessive under some administratively determined
admission control policy. Other admission control criteria, based on
the total resource consumption of a tree may be defined.

7.   QOS-BASED ROUTING AND RESOURCE RESERVATION PROTOCOLS

There must clearly be a well-defined interface between routing and
resource reservation protocols. The nature of this interface, and the
interaction between routing and resource reservation has to be
determined carefully to avoid incompatibilities. The importance of
this can be readily illustrated in the case of RSVP.

RSVP has been designed to operate independent of the underlying
routing scheme. Under this model, RSVP PATH messages establish the
reverse path for RESV messages. In essence, this model is not
compatible with QoS-based routing schemes that compute paths after
receiver reservations are received. While this incompatibility can be
resolved in a simple manner for unicast flows, multicast with
heterogeneous receiver requirements is a more difficult case. For
this, reconciliation between RSVP and QoS-based routing models is
necessary. Such a reconciliation, however, may require some changes
to the RSVP model depending on the QoS-based routing model [ZES97,
ZSSC97, GOA97]. On the other hand, QoS-based routing schemes may be
designed with RSVP compatibility as a necessary goal. How this
affects scalability and other performance measures must be
considered.

8. SECURITY CONSIDERATIONS

Security issues that arise with routing in general are about
maintaining the integrity of the routing protocol in the presence of
unintentional or malicious introduction of information that may lead
to protocol failure [P88]. QoS-based routing requires additional
security measures both to validate QoS requests for flows and to
prevent resource-depletion type of threats that can arise when flows
are allowed to make arbitratry resource requests along various paths
in the network. Excessive resource consumption by an errant flow
results in denial of resources to legitimate flows. While these
situations may be prevented by setting up proper policy constraints,
charging models and policing at various points in the network, the
formalization of such protection requires work [BCCH94].

9. RELATED WORK

"Adaptive" routing, based on network state, has a long history,
especially in circuit-switched networks. Such routing has also been
implemented in early datagram and virtual circuit packet networks.
More recently, this type of routing has been the subject of study in
the context of ATM networks, where the traffic characteristics and
topology are substantially different from those of circuit-switched
networks [MMR96]. It is instructive to review the adaptive routing
methodologies, both to understand the problems encountered and
possible solutions.

Fundamentally, there are two aspects to adaptive, network state-
dependent routing:

   1. Measuring and gathering network state information, and
   2. Computing routes based on the available information.

Depending on how these two steps are implemented, a variety of
routing techniques are possible. These differ in the following
respects:

- what state information is used
- whether local or global state is used
- what triggers the propagation of state information
- whether routes are computed in a distributed or centralized manner
- whether routes are computed on-demand, pre-computed, or in a
  hybrid manner
- what optimization criteria, if any, are used in computing routes
- whether source routing or hop by hop routing is used, and
- how alternate route choices are explored

It should be noted that most of the adaptive routing work has focused
on unicast routing. Multicast routing is one of the areas that would
be prominent with Internet QoS-based routing. We treat this
separately, and the following review considers only unicast routing.
This review is not exhaustive, but gives a brief overview of some of
the approaches.

9.1 Optimization Criteria

The most common optimization criteria used in adaptive routing is
throughput maximization or delay minimization. A general formulation
of the optimization problem is the one in which the network revenue
is maximized, given that there is a cost associated with routing a
flow over a given path [MMR96, K88]. In general, global optimization
solutions are difficult to implement, and they rely on a number of
assumptions on the characteristics of the traffic being routed
[MMR96]. Thus, the practical approach has been to treat the routing
of each flow (VC, circuit or packet stream to a given destination)
independently of the routing of other flows. Many such routing
schemes have been implemented.

9.2 Circuit Switched Networks

Many adaptive routing concepts have been proposed for circuit-
switched networks. An example of a simple adaptive routing scheme is
sequential alternate routing [T88]. This is a hop-by-hop
destination-based routing scheme where only local state information
is utilized. Under this scheme, a routing table is computed for each
node, which lists multiple output link choices for each destination.
When a call set-up request is received by a node, it tries each
output link choice in sequence, until it finds one that can
accommodate the call. Resources are reserved on this link, and the
call set-up is forwarded to the next node. The set-up either reaches
the destination, or is blocked at some node. In the latter case, the

set-up can be cranked back to the previous node or a failure
declared. Crankback allows the previous node to try an alternate
path. The routing table under this scheme can be computed in a
centralized or distributed manner, based only on the topology of the
network. For instance, a k-shortest-path algorithm can be used to
determine k alternate paths from a node with distinct initial links
[T88]. Some mechanism must be implemented during path computation or
call set-up to prevent looping.

Performance studies of this scheme illustrate some of the pitfalls of
alternate routing in general, and crankback in particular [A84, M86,
YS87]. Specifically, alternate routing improves the throughput when
traffic load is relatively light, but adversely affects the
performance when traffic load is heavy. Crankback could further
degrade the performance under these conditions. In general,
uncontrolled alternate routing (with or without crankback) can be
harmful in a heavily utilized network, since circuits tend to be
routed along longer paths thereby utilizing more capacity. This is an
obvious, but important result that applies to QoS-based Internet
routing also.

The problem with alternate routing is that both direct routed (i.e.,
over shortest paths) and alternate routed calls compete for the same
resource. At higher loads, allocating these resources to alternate
routed calls result in the displacement of direct routed calls and
hence the alternate routing of these calls. Therefore, many
approaches have been proposed to limit the flow of alternate routed
calls under high traffic loads. These schemes are designed for the
fully-connected logical topology of long distance telephone networks
(i.e., there is a logical link between every pair of nodes). In this
topology, direct routed calls always traverse a 1-hop path to the
destination and alternate routed calls traverse at most a 2-hop path.

"Trunk reservation" is a scheme whereby on each link a certain
bandwidth is reserved for direct routed calls [MS91]. Alternate
routed calls are allowed on a trunk as long as the remaining trunk
bandwidth is greater than the reserved capacity. Thus, alternate
routed calls cannot totally displace direct routed calls on a trunk.
This strategy has been shown to be very effective in preventing the
adverse effects of alternate routing.

"Dynamic alternate routing" (DAR) is a strategy whereby alternate
routing is controlled by limiting the number of choices, in addition
to trunk reservation [MS91]. Under DAR, the source first attempts to
use the direct link to the destination. When blocked, the source
attempts to alternate route the call via a pre-selected neighbor. If
the call is still blocked, a different neighbor is selected for
alternate routing to this destination in the future. The present call

is dropped. DAR thus requires only local state information. Also, it
"learns" of good alternate paths by random sampling and sticks to
them as long as possible.

More recent circuit-switched routing schemes utilize global state to
select routes for calls. An example is AT&T's Real-Time Network
Routing (RTNR) scheme [ACFH92]. Unlike schemes like DAR, RTNR handles
multiple classes of service, including voice and data at fixed rates.
RTNR utilizes a sophisticated per-class trunk reservation mechanism
with dynamic bandwidth sharing between classes. Also, when alternate
routing a call, RTNR utilizes the loading on all trunks in the
network to select a path. Because of the fully-connected topology,
disseminating status information is simple under RTNR; each node
simply exchanges status information directly with all others.

From the point of view of designing QoS-based Internet routing
schemes, there is much to be learned from circuit-switched routing.
For example, alternate routing and its control, and dynamic resource
sharing among different classes of traffic. It is, however, not
simple to apply some of the results to a general topology network
with heterogeneous multirate traffic. Work in the area of ATM network
routing described next illustrates this.

9.3 ATM Networks

The VC routing problem in ATM networks presents issues similar to
that encountered in circuit-switched networks. Not surprisingly, some
extensions of circuit-switched routing have been proposed. The goal
of these routing schemes is to achieve higher throughput as compared
to traditional shortest-path routing. The flows considered usually
have a single QoS requirement, i.e., bandwidth.

The first idea is to extend alternate routing with trunk reservation
to general topologies [SD95]. Under this scheme, a distance vector
routing protocol is used to build routing tables at each node with
multiple choices of increasing hop count to each destination. A VC
set-up is first routed along the primary ("direct") path. If
sufficient resources are not available along this path, alternate
paths are tried in the order of increasing hop count. A flag in the
VC set-up message indicates primary or alternate routing, and
bandwidth on links along an alternate path is allocated subject to
trunk reservation. The trunk reservation values are determined based
on some assumptions on traffic characteristics. Because the scheme
works only for a single data rate, the practical utility of it is
limited.

The next idea is to import the notion of controlled alternate routing
into traditional link state QoS-based routing [GKR96]. To do this,

first each VC is associated with a maximum permissible routing cost.
This cost can be set based on expected revenues in carrying the VC or
simply based on the length of the shortest path to the destination.
Each link is associated with a metric that increases exponentially
with its utilization. A switch computing a path for a VC simply
determines a least-cost feasible path based on the link metric and
the VC's QoS requirement. The VC is admitted if the cost of the path
is less than or equal to the maximum permissible routing cost. This
routing scheme thus limits the extent of "detour" a VC experiences,
thus preventing excessive resource consumption. This is a practical
scheme and the basic idea can be extended to hierarchical routing.
But the performance of this scheme has not been analyzed thoroughly.
A similar notion of admission control based on the connection route
was also incorporated in a routing scheme presented in [ACG92].

Considering the ATM Forum PNNI protocol [PNNI96], a partial list of
its stated characteristics are as follows:

        o   Scales to very large networks
        o   Supports hierarchical routing
        o   Supports QoS
        o   Uses source routed connection setup
        o   Supports multiple metrics and attributes
        o   Provides dynamic routing

The PNNI specification is sub-divided into two protocols: a signaling
and a routing protocol. The PNNI signaling protocol is used to
establish point-to-point and point to multipoint connections and
supports source routing, crankback and alternate routing. PNNI source
routing allows loop free paths. Also, it allows each implementation
to use its own path computation algorithm. Furthermore, source
routing is expected to support incremental deployment of future
enhancements such as policy routing.

The PNNI routing protocol is a dynamic, hierarchical link state
protocol that propagates topology information by flooding it through
the network. The topology information is the set of resources (e.g.,
nodes, links and addresses) which define the network. Resources are
qualified by defined sets of metrics and attributes (delay, available
bandwidth, jitter, etc.) which are grouped by supported traffic
class. Since some of the metrics used will change frequently, e.g.,
available bandwidth, threshold algorithms are used to determine if
the change in a metric or attribute is significant enough to require
propagation of updated information. Other features include, auto
configuration of the routing hierarchy, connection admission control
(as part of path calculation) and aggregation and summarization of
topology and reachability information.

Despite its functionality, the PNNI routing protocol does not address
the issues of multicast routing, policy routing and control of
alternate routing. A problem in general with link state QoS-based
routing is that of efficient broadcasting of state information. While
flooding is a reasonable choice with static link metrics it may
impact the performance adversely with dynamic metrics.

Finally, Integrated PNNI [I-PNNI] has been designed from the start to
take advantage of the QoS Routing capabilities that are available in
PNNI and integrate them with routing for layer 3. This would provide
an integrated layer 2 and layer 3 routing protocol for networks that
include PNNI in the ATM core. The I-PNNI specification has been
under development in the ATM Forum and, at this time, has not yet
incorporated QoS routing mechanisms for layer 3.

9.4   Packet Networks

Early attempts at adaptive routing in packet networks had the
objective of delay minimization by dynamically adapting to network
congestion. Alternate routing based on k-shortest path tables, with
route selection based on some local measure (e.g., shortest output
queue) has been described [R76, YS81]. The original ARPAnet routing
scheme was a distance vector protocol with delay-based cost metric
[MW77]. Such a scheme was shown to be prone to route oscillations
[B82]. For this and other reasons, a link state delay-based routing
scheme was later developed for the ARPAnet [MRR80]. This scheme
demonstrated a number of techniques such as triggered updates,
flooding, etc., which are being used in OSPF and PNNI routing today.
Although none of these schemes can be called QoS-based routing
schemes, they had features that are relevant to QoS-based routing.

IBM's System Network Architecture (SNA) introduced the concept of
Class of Service (COS)-based routing [A79, GM79]. There were several
classes of service: interactive, batch, and network control. In
addition, users could define other classes. When starting a data
session an application or device would request a COS. Routing would
then map the COS into a statically configured route which marked a
path across the physical network. Since SNA is connection oriented,
a session was set up along this path and the application's or
device's data would traverse this path for the life of the session.
Initially, the service delivered to a session was based on the
network engineering and current state of network congestion. Later,
transmission priority was added to subarea SNA. Transmission
priority allowed more important traffic (e.g. interactive) to proceed
before less time-critical traffic (e.g. batch) and improved link and
network utilization. Transmission priority of a session was based on
its COS.

SNA later evolved to support multiple or alternate paths between
nodes. But, although assisted by network design tools, the network
administrator still had to statically configure routes. IBM later
introduced SNA's Advanced Peer to Peer Networking (APPN) [B85]. APPN
added new features to SNA including dynamic routing based on a link
state database. An application would use COS to indicate it traffic
requirements and APPN would calculate a path capable of meeting these
requirements. Each COS was mapped to a table of acceptable metrics
and parameters that qualified the nodes and links contained in the
APPN topology Database. Metrics and parameters used as part of the
APPN route calculation include, but are not limited to: delay, cost
per minute, node congestion and security. The dynamic nature of APPN
allowed it to route around failures and reduce network configuration.

The service delivered by APPN was still based on the network
engineering, transmission priority and network congestion. IBM later
introduced an extension to APPN, High Performance Routing
(HPR)[IBM97]. HPR uses a congestion avoidance algorithm called
adaptive rate based (ARB) congestion control. Using predictive
feedback methods, the ARB algorithm prevents congestion and improves
network utilization. Most recently, an extension to the COS table
has been defined so that HPR routing could recognize and take
advantage of ATM QoS capabilities.

Considering IP routing, both IDRP [R92] and OSPF support type of
service (TOS)-based routing. While the IP header has a TOS field,
there is no standardized way of utilizing it for TOS specification
and routing. It seems possible to make use of the IP TOS feature,
along with TOS-based routing and proper network engineering, to do
QoS-based routing. The emerging differentiated services model is
generating renewed interest in TOS support. Among the newer schemes,
Source Demand Routing (SDR) [ELRV96] allows on-demand path
computation by routers and the implementation of strict and loose
source routing. The Nimrod architecture [CCM96] has a number of
concepts built in to handle scalability and specialized path
computation. Recently, some work has been done on QoS-based routing
schemes for the integrated services Internet. For example, in [M98],
heuristic schemes for efficient routing of flows with bandwidth
and/or delay constraints is described and evaluated.

9. SUMMARY AND CONCLUSIONS

In this document, a framework for QoS-based Internet routing was
defined. This framework adopts the traditional separation between
intra and interdomain routing. This approach is especially meaningful
in the case of QoS-based routing, since there are many views on how
QoS-based routing should be accomplished and many different needs.
The objective of this document was to encourage the development of

different solution approaches for intradomain routing, subject to
some broad requirements, while consensus on interdomain routing is
achieved. To this end, the QoS-based routing issues were described,
and some broad intradomain routing requirements and an interdomain
routing model were defined. In addition, QoS-based multicast routing
was discussed and a detailed review of related work was presented.

The deployment of QoS-based routing across multiple administrative
domains requires both the development of intradomain routing schemes
and a standard way for them to interact via a well-defined
interdomain routing mechanism. This document, while outlining the
issues that must be addressed, did not engage in the specification of
the actual features of the interdomain routing scheme. This would be
the next step in the evolution of wide-area, multidomain QoS-based
routing.

REFERENCES

    [A79]   V. Ahuja, "Routing and Flow Control in SNA", IBM Systems
             Journal, 18 No. 2, pp. 298-314, 1979.

    [A84]   J. M. Akinpelu, "The Overload Performance of Engineered
             Networks with Non-Hierarchical Routing", AT&T Technical
             Journal, Vol. 63, pp. 1261-1281, 1984.

    [ACFH92] G. R. Ash, J. S. Chen, A. E. Frey and B. D. Huang, "RealTime
             Network Routing in a Dynamic Class-of-Service Network",
             Proceedings of ITC 13, 1992.

    [ACG92] H. Ahmadi, J. Chen, and R. Guerin, "Dynamic Routing and Call
             Control in High-Speed Integrated Networks", Proceedings of
             ITC-13, pp. 397-403, 1992.

    [B82]   D. P. Bertsekas, "Dynamic Behavior of Shortest Path Routing
             Algorithms for Communication Networks", IEEE Trans. Auto.
             Control, pp. 60-74, 1982.

    [B85]   A. E. Baratz, "SNA Networks of Small Systems", IEEE JSAC,
             May, 1985.

    [BBCD98] Black, D., Blake, S., Carlson, M., Davies, E., Wang, Z., and
             W. Weiss, "An Architecture for Differentiated Services",
             Work in Progress.

    [BCCH94] Braden, R., Clark, D., Crocker, D., and C. Huitema, "Report
             of IAB Workshop on Security in the Internet Architecture",
             RFC 1636, June 1994.

[BCF94]   A. Ballardie, J. Crowcroft and P. Francis, "Core-Based
          Trees: A Scalable Multicast Routing Protocol", Proceedings
          of SIGCOMM '94.

[BCS94]   Braden, R., Clark, D., and S. Shenker, "Integrated Services
          in the Internet Architecture: An Overview", RFC 1633, July
          1994.

[BZ92]   S. Bahk and M. El Zarki, "Dynamic Multi-Path Routing and How
          it Compares with Other Dynamic Routing Algorithms for High
          Speed Wide Area Networks", Proc. SIGCOMM '92, pp. 53-64,
          1992.

[BZBH97] Braden, R., Zhang, L., Berson, S., Herzog, S., and S. Jamin,
          "Resource ReSerVation Protocol (RSVP) -- Version 1
          Functional Spec", RFC 2205, September 1997.

[C91]   C-H. Chow, "On Multicast Path Finding Algorithms",
          Proceedings of the IEEE INFOCOM '91, pp. 1274-1283, 1991.

[CCM96]   Castineyra, I., Chiappa, J., and M. Steenstrup, "The Nimrod
          Routing Architecture", RFC 1992, August 1996.

[DEFV94] S. E. Deering, D. Estrin, D. Farinnacci, V. Jacobson, C-G.
          Liu, and L. Wei, "An Architecture for Wide-Area Multicast
          Routing", Technical Report, 94-565, ISI, University of
          Southern California, 1994.

[ELRV96] Estrin, D., Li, T., Rekhter, Y., Varadhan, K., and D.
          Zappala, "Source Demand Routing: Packet Format and
          Forwarding Specification (Version 1)", RFC 1940, May 1996.

[GKR96]   R. Gawlick, C. R. Kalmanek, and K. G. Ramakrishnan, "On-Line
          Routing of Permanent Virtual Circuits", Computer
          Communications, March, 1996.

[GPSS98] A. Ghanwani, J. W. Pace, V. Srinivasan, A. Smith and M.
          Seaman, "A Framework for Providing Integrated Services over
          Shared and Switched IEEE 802 LAN Technologies", Work in
          Progress.

[GM79]   J. P. Gray, T. B. McNeil, "SNA Multi-System Networking", IBM
          Systems Journal, 18 No. 2, pp. 263-297, 1979.

[GOA97]   Y. Goto, M. Ohta and K. Araki, "Path QoS Collection for
          Stable Hop-by-Hop QoS Routing", Proc. INET '97, June, 1997.

   [GKOP98] R. Guerin, S. Kamat, A. Orda, T. Przygienda, and D.
            Williams, "QoS Routing Mechanisms and OSPF extensions", work
            in progress, March, 1998.

   [IBM97] IBM Corp, SNA APPN - High Performance Routing Architecture
            Reference, Version 2.0, SV40-1018, February 1997.

   [IPNNI] ATM Forum Technical Committee. Integrated PNNI (I-PNNI) v1.0
            Specification. af-96-0987r1, September 1996.

   [ISI81] Postel, J., "Internet Protocol", STD 5, RFC 791, September
            1981.

   [JMW83] J. M. Jaffe, F. H. Moss, R. A. Weingarten, "SNA Routing:
            Past, Present, and Possible Future", IBM Systems Journal,
            pp. 417-435, 1983.

   [K88]   F.P. Kelly, "Routing in Circuit-Switched Networks:
            Optimization, Shadow Prices and Decentralization", Adv.
            Applied Prob., pp. 112-144, March, 1988.

   [L95]   W. C. Lee, "Topology Aggregation for Hierarchical Routing in
            ATM Networks", ACM SIGCOMM Computer Communication Review,
            1995.

   [M86]   L. G. Mason, "On the Stability of Circuit-Switched Networks
            with Non-hierarchical Routing", Proc. 25th Conf. On Decision
            and Control, pp. 1345-1347, 1986.

   [M98]   Moy, J., "OSPF Version 2", STD 54, RFC 2328, April 1998.

   [M94]   Moy, J., "MOSPF: Analysis and Experience", RFC 1585, March
            1994.

   [M98]   Q. Ma, "Quality-of-Service Routing in Integrated Services
            Networks", PhD thesis, Computer Science Department, Carnegie
            Mellon University, 1998.

   [MMR96] D. Mitra, J. Morrison, and K. G. Ramakrishnan, "ATM Network
            Design and Optimization: A Multirate Loss Network
            Framework", Proceedings of IEEE INFOCOM '96, 1996.

   [MRR80] J. M. McQuillan, I. Richer and E. C. Rosen, "The New Routing
            Algorithm for the ARPANET", IEEE Trans. Communications, pp.
            711-719, May, 1980.

   [MS91]   D. Mitra and J. B. Seery, "Comparative Evaluations of
             Randomized and Dynamic Routing Strategies for Circuit
             Switched Networks", IEEE Trans. on Communications, pp. 102-
             116, January, 1991.

   [MW77]   J. M. McQuillan and D. C. Walden, "The ARPANET Design
             Decisions", Computer Networks, August, 1977.

   [NC94]   Nair, R. and Clemmensen, D. : "Routing in Integrated
             Services Networks", Proc. 2nd International Conference on
             Telecom. Systems Modeling and Analysis, March 1994.

   [P88]   R. Perlman, "Network Layer Protocol with Byzantine
             Robustness", Ph.D. Thesis, Dept. of EE and CS, MIT, August,
             1988.

   [PNNI96] ATM Forum PNNI subworking group, "Private Network-Network
             Interface Spec. v1.0 (PNNI 1.0)", afpnni-0055.00, March
             1996.

   [R76]   H. Rudin, "On Routing and "Delta Routing": A Taxonomy and
             Performance Comparison of Techniques for Packet-Switched
             Networks", IEEE Trans. Communications, pp. 43-59, January,
             1996.

   [R92]   Y. Rekhter, "IDRP Protocol Analysis: Storage Overhead", ACM
             Comp. Comm. Review, April, 1992.

   [R96]   B. Rajagopalan, "Efficient Link State Routing", Work in
             Progress, available from braja@ccrl.nj.nec.com.

   [RN98]   B. Rajagopalan and R. Nair, "Multicast Routing with Resource
             Reservation", to appear in J. of High Speed Networks, 1998.

   [SD95]   S. Sibal and A. Desimone, "Controlling Alternate Routing in
             General-Mesh Packet Flow Networks", Proceedings of ACM
             SIGCOMM, 1995.

   [SPG97]   Shenker, S., Partridge, C., and R. Guerin, "Specification of
             Guaranteed Quality of Service", RFC 2212, September 1997.

   [T88]   D. M. Topkis, "A k-Shortest-Path Algorithm for Adaptive
             Routing in Communications Networks", IEEE Trans.
             Communications, pp. 855-859, July, 1988.

   [W88]   B. M. Waxman, "Routing of Multipoint Connections", IEEE
             JSAC, pp. 1617-1622, December, 1988.

   [W97]   Wroclawski, J., "Specification of the Controlled-Load Network
            Element Service", RFC 2211, September 1997.

   [WC96]   Z. Wang and J. Crowcroft, "QoS Routing for Supporting
            Resource Reservation", IEEE JSAC, September, 1996.

   [YS81]   T. P. Yum and M. Schwartz, "The Join-Based Queue Rule and
            its Application to Routing in Computer Communications
            Networks", IEEE Trans. Communications, pp. 505-511, 1981.

   [YS87]   T. G. Yum and M. Schwartz, "Comparison of Routing Procedures
            for Circuit-Switched Traffic in Nonhierarchical Networks",
            IEEE Trans. Communications, pp. 535-544, May, 1987.

   [ZES97] Zappala, D., Estrin, D., and S. Shenker, "Alternate Path
            Routing and Pinning for Interdomain Multicast Routing", USC
            Computer Science Technical Report #97-655, USC, 1997.

   [ZSSC97] Zhang, Z., Sanchez, C., Salkewicz, B., and E. Crawley, "QoS
            Extensions to OSPF", Work in Progress.

AUTHORS' ADDRESSES

    Bala Rajagopalan
    NEC USA, C&C Research Labs
    4 Independence Way
    Princeton, NJ 08540
    U.S.A

    Phone: +1-609-951-2969
    EMail: braja@ccrl.nj.nec.com

    Raj Nair
    Arrowpoint
    235 Littleton Rd.
    Westford, MA 01886
    U.S.A

    Phone: +1-508-692-5875, x29
    EMail: nair@arrowpoint.com

    Hal Sandick
    Bay Networks, Inc.
    1009 Slater Rd., Suite 220
    Durham, NC 27703
    U.S.A

    Phone: +1-919-941-1739
    EMail: Hsandick@baynetworks.com

    Eric S. Crawley
    Argon Networks, Inc.
    25 Porter Rd.
    Littelton, MA 01460
    U.S.A

    Phone: +1-508-486-0665
    EMail: esc@argon.com