Network Working Group
Request for Comments: 3387
Category: Informational

M. Eder
H. Chaskar
Nokia
S. Nag

September 2002

Considerations from the Service Management Research Group (SMRG) on Quality of Service (QoS) in the IP Network

Status of this Memo

This memo provides information for the Internet community. It does not specify an Internet standard of any kind. Distribution of this memo is unlimited.

Copyright Notice

Abstract

The guiding principles in the design of IP network management were simplicity and no centralized control. The best effort service paradigm was a result of the original management principles and the other way around. New methods to distinguish the service given to one set of packets or flows relative to another are well underway. However, as IP networks evolve the management approach of the past may not apply to the Quality of Service (QoS)-capable network envisioned by some for the future. This document examines some of the areas of impact that QoS is likely to have on management and look at some questions that remain to be addressed.

1. Introduction

Simplicity above all else was one of the guiding principles in the design of IP networks. However, as IP networks evolve, the concept of service in IP is also evolving, and the strategies of the past may not apply to the full-service QoS-capable network envisioned by some for the future. Within the IP community, their exists a good deal of impetus for the argument that if the promise of IP is to be fulfilled, networks will need to offer an increasing variety of services. The definition of these new services in IP has resulted in a need for reassessment of the current control mechanism utilized by IP networks. Efforts to provide mechanisms to distinguish the service given to one set of packets or flows relative to another are well underway, yet many of the support functions necessary to exploit these mechanisms are limited in scope and a complete framework is

non-existent. This is complicated by the fact that many of these new services will also demand some form of billing framework in addition to a control one, something radically new for IP.

This document intends to evaluate the network and service management issues that will need to be addressed, if the IP networks of the future are going to offer more than just the traditional best effort service in any kind of significant way.

2. Background

The task of defining a management framework for QoS will be difficult due to the fact that it represents a radical departure from the best effort service model that was at the core of IP in the past, and had a clear design strategy to have simplicity take precedence over everything else [1]. This philosophy was nowhere more apparent than in the network and service management area for IP [2]. Proposed changes to support a variety of QoS features will impact the existing control structure in a very dramatic way. Compounding the problem is the lack of understanding of what makes up a "service" in IP [3]. Unlike some other network technologies, in IP it does not suffice to limit the scope of service management simply to end-to-end connectivity, but the transport service offered to packets and the way the transport is used must also be covered. QoS management is a subset of the more general service management. In looking to solve the QoS management problem it can be useful to understand some of the issues and limitations of the service management problem. QoS can not be treated as a standalone entity and will have its management requirements driven by the general higher level service requirements. If the available transport services in IP expand, the result will be the further expansion of what is considered a service. The now de-facto inclusion of WEB services in the scope of IP service, which is remarkable given that the WEB did not even exist when IP was first invented, illustrates this situation well. This phenomenon can be expected to increase with the current trend towards moving network decision points towards the boundary of the network and, as a result, closer to the applications and customers. Additionally, the argument continues over the need for QoS in IP networks at all. New technologies based on fiber and wavelength-division multiplexing have many people convinced that bandwidth will be so inexpensive it is not going to be necessary to have an explicit control framework for providing QoS differentiation. However uneconomical it is to engineer a network for peak usage, a major argument in this debate certainly is the cost of developing operational support systems for a QoS network and deploying them in the existing networks. Just the fact that customers might be willing to pay for additional service may not be justification for implementing sweeping architectural changes that could seriously affect the Internet as it is known

today. The IP community must be very concerned that the equality that characterized the best effort Internet may be sacrificed in favor of a service that has a completely different business model. If the core network started to provide services that generated more revenue, it could easily come at the expense of the less revenue generating best effort service.

3. IP Management Standardization

Management standardization efforts in the IP community have traditionally been concerned with what is commonly referred to as "element management" or "device management". Recently, new efforts in IP management have added the ability to address service issues and to look at the network in more abstract terms. These efforts which included a logical representation of services as well as the representation of resources in the network, combined with the notion of a user of a service, has made possible the much talked about concept of 'policy'. Notable among these efforts are the Policy work in the IETF and the DMTF work on CIM and DEN. Crucial elements of the service management framework are coming into perspective, but point to a trend in IP that is a quite radical departure from the control mechanisms of the past. As the service model evolves from being what was sufficient to support best effort to being able to support variable levels of service, a trend towards a centralized management architecture has become quite apparent.

This is becoming increasingly apparent for two reasons. QoS mechanisms need network wide information [4], and for them to succeed, they must not require a tremendous amount of support from the core network. It is becoming increasingly accepted that only at the edge of the network will there be sufficient resources to provide the mechanisms necessary to admit and control various QoS flows.

A question often asked these days is if "the architectural benefits of providing services in the middle of the network outweigh the architectural costs"[5]. This same question should be asked of service management. As new network elements are needed to support service management, even if they are not contributing directly to the forwarding of packets, the cost both in the increased complexity and the possibility of destabilizing the networks needs to be considered. An analyses of this issue will be made by the SMRG when we start to look more in detail at some of the issues raised in this survey document.

4. Telecommunications Service Management

One place to start an effort to define service management in IP networks is by looking at what has been done previously in telecommunications networks. The telecommunications standards for a service management framework have not received wide scale acceptance even in an environment in which the service is fairly constrained. Many proprietary protocols still dominate in the market even though regulation has made it necessary for network operators to open their networks sufficiently to allow for multiple vendor participation in providing the service. This indicates that some formalized boundaries exist or the markets are sufficiently large to justify the development of interfaces. International telecommunications management standards look at the complete management problem by dividing it into separate but highly related layers. Much of the terminology used to describe the management problem in IP has diffused from the telecommunications standards [6]. These standards were designed specifically to address telecommunications networks and services, and it is not clear how applicable they will be to IP networks. Service management is defined in terms of the set of services found in telecommunications networks and the management framework reflects the hierarchical centralized control structure of these networks. The framework for service management is based on the Telecommunications Management Network (TMN) layered approach to management. Current IP standards are heavily weighted towards the element management layer and especially towards the gathering of statistical data with a decentralized approach being emphasized. In the TMN architecture a dependency exists between layers and clear interfaces at the boundaries are defined. To what extent service management, as defined in the TMN standards, can be applied to IP where there would likely be resistance to a requirement to have formalized interfaces between layers [6] must be further investigated.

TMN concepts must be applied carefully to IP networks because fundamental differences exist. Control of IP networks is highly distributed especially in the network layer. Management is non-hierarchical and decentralized with many peer-to-peer relationships. A formal division of management into layers, where management dependencies exist at the borders of these layers, may not be applicable to IP. Any effort to define service management in IP must be constantly vigilant that it does not assume the telecommunications concepts can be applied directly to IP networks. The most basic abstraction of the network management problem into element, network, and service management has its origins in the telecommunications industry's standardization work and the IP management framework might not have made even these distinctions if it where not for the telecommunications legacy.

5. IP Service Management: Problem Statement

In defining the Service Management Framework for IP, the nature of services that are going to need to be managed must be addressed. Traditionally network management frameworks consist of two parts, an informational framework and the framework to distribute information to the network devices. A very straight forward relationship exists in that the distribution framework must support the informational one, but also more subtle relationships exists with what the informational and distribution frameworks imply about the management of the system. The informational framework appears to be the easier problem to address and the one that is principally being focused on by the IP community.

Efforts like the DMTF CIM are currently trying to define network, and to a lesser extent service, information models. These efforts show a surprising similarity to those of the telecommunications industry to define information models [7]. What has not emerged is a standard for defining how the information contained in the models is to be used to manage a network.

The number of elements to be managed in these networks will require this information to be highly distributed. Highly distributed directories would be a prime candidate for the information that is of a static nature. For information that is of a dynamic nature the problem becomes far more complex and has yet to be satisfactorily addressed. Policy management is a logical extension of having distributed directories services available in the network. The IETF and DMTF are looking to Policy management to be a framework to handle certain service management issues. Much of the current policy efforts are focused on access and traffic prioritization within a particular network element and only for a single administrative domain [8]. Classifying traffic flows and enforcing policies at the edge with the intent of focusing on admission issues, without addressing the end-to-end nature of the problem, leaves some of the most complex QoS management issues still unanswered. Providing a verifiable commodity level of service, in IP, will effect every facet of the network and a management solution to the problem will have to address the scale and the dynamics by which it operates.

5.1 Common Management Domain

Standardization efforts need to concentrate on the management problems that are multi-domain in character. The test for multi-domain often centers around there being a many-to-one or a one-to-many relationship requiring the involvement of two or more distinct entities. Domains could reflect the administrative domain, routing domain, or include agreements between domains. Unlike the

telecommunications network in which traffic traverses only a relatively small number of domains, traffic in IP networks is likely to traverse numerous domains under separate administrative control. Further complicating the situation is, that unlike the telecommunications network, many of these domains will be highly competitive in nature, offering and accommodating varying service level agreements. Telecommunications traffic, even with deregulation, passes from the access providers network to a core network and then, if it is an international call, across international boundaries. The number of domains is relative to IP small, the service supported in each is virtually identical, and yet each domains is likely to have a different business model from the other. In contrast IP will have many domains, many services, and domains will likely be highly competitive. To be successful IP will need to model the domain problem in a way that reduces the complexity that arises from having many independent networks each having a different service model being responsible for a single flow. Addressing service management issues across domains that are direct competitors of each other will also complicate the process because a solution must not expose too much information about the capabilities of one domains network to the competitor. Solutions may require a 3rd party trusted by both to provide the needed management functions while at the same time insuring that sensitive information does not pass from one to the other.

5.2 Service Management Business Processes

A service management framework must address the business processes that operate when providing a service. A service can be separated into two fundamental divisions. The first is the definition of the service and the second is the embodiment of the service. While this division may seem intuitive, a formal process that addresses these two aspects of a service needs to be in place if management of the service is to be actually realized.

In specifying a service it must be possible to map it onto the capabilities of the underlying network architecture. The service needs to be specified in an unambiguous way so that mechanisms can be put in place to enable the control of the service. It can be a useful tool to view the relationship of the definition of a service to an instance of that service to the relationship between the definition of an object to the instantiation of that object in object oriented modeling. As networks evolve it is going to be necessary to logically describe the network capabilities to the service and because IP networks are so fragmented specific service classifications will need to be made available that transcend the individual regions and domains. An interface that defines and

controls the network capabilities, abstracted for the service perspective, allows for the administration of the network by the service management systems.

Services are often designed with management capabilities specific to them. These services have tended to not rely on the service aspects of the network, but only on its transport capabilities. As services become more dependent on the network, Management over a shared framework will be required. Operators have recognized the business need to allow the user to have as much control over the management of their own services as possible. IP services will be highly diverse and customizable further necessitating that the management of the service be made available to the user to the extent possible.

In the IP environment where they may be many separate entities required to provide the service this will create a significant management challenge.

5.3 Billing and Security

Paramount to the success of any service is determining how that service will be billed. The process by which billing will take place must be defined at the service inception. It is here that the network support necessary for billing should be addressed. Analogously, security must also be addressed in the most early stages of the service definition. It is not practical to assume that the billing and the security services will be hosted by the same provider as the service itself or that it will be possible to have the billing and security functions specifically designed for every service. These functions will have to be a generic part of the network.

5.4 Standards

Given the limited success of the telecommunications standards bodies efforts to formalize the relationship between different management support functions it is highly suspect that such efforts would succeed in IP networks which have an even more diverse concept of network and services. If the IP network is to be made up of peer domains of equal dominion it will be necessary to have management functionality that is able to traverse these domains. Of course the perspective of where management responsibility lies is largely dependent on the reference point. A centric vantage point indicates responsibility shared equally among different domains. From within any particular domain management responsibility exists within that domain and that domain only. For a management framework to succeed in IP networks logical management functions will have to be identified along with an extremely flexible definition language to define the interface to these management functions. The more the

management functionality will have to cross boundaries of responsibility, the more the network management functions have to be distributed throughout the network.

5.5 Core Inter-domain Functions

The service management paradigm for IP must address management from a perspective that is a combination of technical solutions as well as a formula for representing vendor business relationships. Currently services that need support between domains require that the service level agreements (SLAs) be negotiated between the providers. At some point these agreements will likely become unmanageable, if the number of agreements becomes very large and/or the nature of the agreements is highly variable. This will result in there being sufficient need for some form of standardization to control these agreements.

Bandwidth Brokers have been conceived as a method for dealing with many of the problems between the domains relating to traffic from a business perspective. The premise of the Bandwidth Brokers is to insure agreement between the network domains with regards to traffic, but security and billing issues, that are not likely to be as quantifiable, will also need to be addressed. Service providers have traditionally been reluctant to use bandwidth broker or SLA types of functions as they fear such tools expose their weaknesses to competitors and customers. While this is not a technical problem, it does pose a real practical problem in managing a service effectively. Looking at the basic requirements of the QoS network of the future two competing philosophies become apparent. The network providers are interested in having more control over the traffic to allow them to choose what traffic gets priority especially in a congested environment. Users desire the ability to identify a path that has the characteristics very similar to a leased line [9]. In either situation as IP bandwidth goes from being delivered on an equal basis, to being delivered based on complex formulas, there will become an increasing need to provide authentication and validation to verify who gets what service and that they pay for it. This will include the ability to measure that the service specified is being provided, to define the exact parameters of the service, and to verify that only an authorized level of service is being provided.

Some of the earlier work on an architectural framework for mixed traffic networks has suggested that bilateral agreements will be the only method that will work between administrative domains [10]. Multilateral agreements may indeed be complex to administer, but bilateral agreements will not scale well and if the traffic needs to traverse many administrative domains it will be hard to quantify the

end-to-end service being provided. Instability in the ownership and administration of domains will also limit the usability of bilateral agreements in predicting end-to-end service.

As the convergence towards all IP continues it will be interesting to understand what effects existing telecommunications regulations might have on IP networks as more regulated traffic is carried over them. Regulation has been used in the telecommunications world to open the network, but it has had mixed results. A regulated process could possibly eliminate the effects competitive pressures will have on bilateral types of agreements and make it possible to get a truly open environment, but it could also have an opposite effect. Unfortunately the answer to this question may not come in the form of the best technical solution but in the politically most acceptable one. If traffic agreements between the boundaries of networks is not standardized a continuing consolidation of network providers would result. Providers unable to induce other providers to pair with them may not be able to compete if QoS networks become commonplace. This would be especially visible for small and midsize service providers, who would be pressured to combine with a larger provider or face not being able to offer the highest levels of service. If this phenomenon plays out across international boundaries it is hard to predict what the final outcome might be.

5.6 Network Services

The majority of current activity on higher level management functions for IP networks have been restricted to the issue of providing QoS. Many service issues still remain to be resolved with respect to the current best effort paradigm and many more can be expected if true QoS support is realized. Authentication, authorization and accounting services still inadequate for the existing best effort service will need additional work to support QoS services.

It is reasonable that services can be classified into application level services and transport level services. Transport services are the services that the network provides independent of any application. These include services such as Packet Forwarding and Routing, QoS differentiation, Traffic Engineering etc. These might also include such functions as security (Ipsec) and Directory services. In IP networks a distinction is often made between QoS transport services that are viewed as end-to-end (RSVP) or per-hop (Diffserv). From a management perspective the two are very similar. Transport level services are not very flexible, requiring application level services to fit into the transport framework. An application that needs additional transport level services will need to be a mass-market application where the investment in new infrastructure can be justified. Because of the effort in altering transport

services, applications that need new ones will have a longer time to market and the effort and cost to develop a framework necessary to support new transport services should not be underestimated.

Application level services are those specific to the application. Many service management functions occur between the application supplier and the application consumer which require no knowledge or support by the existing network. By keeping service management functions at this level time to market and costs can be greatly reduced. The disadvantages are that many applications need the same functionality causing inefficient use of the network resources. Services supplied by the network are able to be built more robustly and can provide additional functionality, by virtue of having access to information that applications can not, providing additional benefit over application level services. An example of an application level service that could benefit from a Network service is the AAA paradigm for Web based E-Commerce, which is largely restricted to user input of credit card information. Sometimes application level service requirements have the disadvantages of both transport service and application service level. For instance, in IP telephony, this may include services provided by a gateway or other network device specific to IP telephony to support such services as call forwarding or call waiting. The mass appeal of IP telephony makes it possible to suggest considerable infrastructure changes, but the nature of this kind of change has contributed to the slow penetration of IP telephony applications.

6. The Way to a QoS Management Architecture

An overview of some of the problems in the previous sections shows a need for a consolidated framework. Transport level QoS will demand traffic engineering that has a view of the complete network that is far more comprehensive than what is currently available via the Routing protocols. This view will need to including dynamic network congestion information as well as connectivity information. The current existing best-effort transport control may become more of a hindrance to new services and may be of questionable value if the IP network will truly become a full service QoS network. Both IntServ and DiffServ QoS schemes require network provisioning to adequately support QoS within a particular domain and agreements for traffic traversing domains. Policy management, object oriented information models, and domain gateways are leading to a more centralized management structure that provides full service across domains and throughout the network. Given the probable cost and complexity of such a system failure to come up with a standard, even if it is a de facto one, will have serious implications for the Internet in the future.

6.1 Point to Point QoS

For the current trends in QoS to succeed, there will need to be harmonization across the new and existing control structures. By utilizing a structure very similar to the existing routing control structures, it should be possible develop functionality, not in the data path, that can allocate traffic within a domain and use inter-domain signaling to distribute between domains. Additional functionality, necessary to support QoS-like authorization and authentication functions for edge devices admitting QoS traffic and administering and allocating traffic between administrative domains could also be supported. While meeting the requirements for a bandwidth broker network element [10], additional functionality of making more general policy decisions and QoS routing could also be performed. Given that these tasks are interrelated it makes sense to integrate them if possible.

The new service architecture must allocate traffic within a particular administrative domain and signal traffic requirements across domains, while at the same time it must be compatible with the current method for routing traffic. This could be accomplished by redirecting routing messages to a central function, which would then calculate paths based on the entire network transport requirements. Across domains, communication would occur as necessary to establish and maintain service levels at the gateways. At the edges, devices would provide traffic information to billing interfaces and verify that the service level agreed to was being provided. For scalability any central function would need to be able to be distributed in large networks. Routing messages, very similar in content to the existing ones, would provide information sufficient to support the traffic engineering requirements without changing the basic forwarding functions of the devices. Having routes computed centrally would simplify network devices by alleviating them from performing computationally intensive routing related tasks.

Given the number of flows through the network the core can not know about individual flow states [11]. At the same time it is not practical to expect that the edge devices can determine paths that will optimally utilize the network resources. As the information needed to forward traffic through the network becomes related to complex parameters that can not be determined on a per hop basis and have nothing to do with the forwarding of packets, which routers do best, it might make sense to move the function of determining routes to network components specifically designed for the task. In a QoS network routing decisions will become increasingly dependent on information not easily discernable from the data that routers could logically share between themselves. This will necessitate the need to for additional functionality to determine the routing of data

through the network and further suggests that all the information needed to allow a router to forward packets might not be better provided by a network element external to the packet forwarding functions of a router.

At the edges of the network where the traffic is admitted it will be necessary to have mechanisms that will insure the traffic is within the bounds of what has been specified. To achieve this it will be necessary to buffer and control the input traffic. Second the traffic would need to be marked so the other network elements are able to identify that this is preferred traffic without having to keep flow information. Conversely, a path could be chosen for the traffic that was dedicated to the level of service being requested that was per flow based. A combination of the two would be possible that would allow a reservation of resources that would accommodate multiple flows. Both methods are similar from a management perspective and are really identical with regards to route determination that could be performed centrally in that one method represents just a virtual path based on the handling of the packets by the device in the network and the second would be a pre-reserved path through the network. Existing best effort routing will not provide the optimum routes for these new levels of service and to achieve this it would be necessary to have either routing protocols that supported optimum path discovery or mechanisms to configure paths necessary to support the required services. In addition to specific service parameters reliability will also be a potential service discriminator. It is unlikely using traditional path determination methods that in the event of a failure a new path could be determined sufficiently quickly to maintain the agreed service level. This would imply the need for multiple path reservations in some instances. Because Per flow reservations are too resource intensive virtual trunks would provide a good way to reduce the amount of management traffic by reserving blocks of capacity and would provide stability in the event of a failure in the resource reservation and route selection functions.

There are implications of providing shaping at the network boundaries. Shaping would include both rate and burst parameters as well as possible delay aspects. Having to provision services with specific service parameters would present both major business and technical problems. By definition, packet data is bursty in nature and there exist periods of idleness during the session that a provider could reasonable hope to exploit to better utilize the network resources. It is not practical to expect a consumer paying a premium for a service would not check that the service was truly available. Such a service model seems to be filled with peril for

the existing best effort Internet, because any significant amount of bandwidth that was reserved for exclusive use or a high priority flow would not be available for best effort data.

With respect to traffic within the network itself there will be the need to pre-configure routes and to provide the ability to have routes be dynamically configured. Some of the problems with pre-configured traffic include the basic inconsistency with the way traffic is currently engineered through the Internet and the difficulty in developing arrangements between administrative domains. The current Internet has been developed with one of the most egalitarian yet simplistic methods of sharing bandwidth. Supporting the existing best effort service, in an unbiased way, while at the same time providing for other classes of service could potentially add a tremendous amount of complexity to the QoS scheme. On the other hand, if the reserved bandwidth is not shared it could result in a significant impact on the availability of the bandwidth in the Internet as we know it today. QoS could potentially contribute more to their being insufficient bandwidth, by reserving bandwidth within the network that can not be used by other services, even though it can be expected that this bandwidth will be underutilized for much of the time. Add to that the motivation of the service providers in wanting to sell commodity bandwidth, and there could be tremendous pressures on the availability of Internet bandwidth.

Current work within the IP community on defining mechanisms to provide QoS have centered on a particular few architectures and a handful of new protocols. In the following sections, we will examine some of the particular issues with regards to the current IP community efforts as they relate to the previous discussions. It is not the goal of this document to serve as a tutorial on these efforts but rather to identify some of the support issues related to using particular technologies that support some form of classifiable service within an IP network.

6.2 QoS Service Management Scope

One can restrict the scope of a discussion of QoS management only to the configuration of a path between two endpoints. Even within this limited scope there still remains many unresolved issues. There is no expectation that a QoS path for traffic between two points needs to be, or should be, the same in both directions. Given that there will be an originator of the connection there are questions about how billing and accounting with be resolved if the return path is established by a different provider then that of the originator of the connection. To facilitate billing a method will need to exist that permits the application originating the call to pay also for the return path and also for collect calls to be made. 3rd party

providers will need to be established that are trusted by all parties in the data path to insure billing and guaranteed payment. Utilizing the service of a virtual DCN that is built upon both IETF and non-IETF protocols, messages between service providers and the 3rd party verification system can be secured. A signaling protocol will be necessary to establish the cost of the call and who will be paying for it, and each provider will need a verifiable method to bill for the service provided. As pointed out earlier this functionality will be similar to what is used in the existing telephone network, but will be at a much larger scale and potentially involve providers that are highly competitive with each other.

7. The DiffServ Architecture

The DiffServ management problem is two pronged. First there is the management within the administrative domain that must be addressed, and then the management between the domains. There has been little actual work on the second in the architecture. What work there has been anticipates that service level agreements will be reached between the administrative domains, and that end-to-end service will be a concatenation of these various service level agreements. This is problematic for many reasons. It presumes that agreements reached bilaterally could be concatenated and continue to provide a level of end-to-end service the customer would be willing to pay a premium for. Problems discussed earlier, with trying to maintain large numbers of these agreements between competitive networks would also apply, and tend to limit the effectiveness of this approach. To efficiently establish the chain necessary to get end to end service it might take an infinite number of iterations.

Guaranteeing a class of service on a per hop basis is in no way a guarantee of the service on an end-to-end basis. It is not likely that a customer would be willing to pay for an improved level of service if it did not include guarantees on the bandwidth and the quantitative bounds on delay and error rates guaranteed end-to-end. This would necessitate engineering the paths through the network so as to achieve a desired end-to-end result. While it is very likely that an initial attempt at providing this kind of service will specify only a particular ingress and egress border, for robustness and flexibility it will be desirable to have a network that can support such service without such limitations. The Intserv approach, as opposed to the DiffServ architecture, would require per flow information in the core network and may as a result of this prove not to be scalable [11]. A DiffServ type architecture, with a limited number of service classes, could be pre-provisioned, and as network circumstances warranted, be modified to support the actual dynamics of the network.

The high level functional requirements for edge routers has been quite well defined in the DiffServ architecture, but the true scope of the effort to implement this functionality has not been well recognized. While interesting differences exist between the QoS architecture of the Internet and the circuit switched network used for telecommunications much of the lessons learned in telecommunications should, even if they might do little else, provide some insight into the level of effort needed to implement these kinds of requirements. Ironically, given the Internet community in the past has rejected the level of standardization that was proposed for management of telecommunications networks, it may be the full service internet where it becomes actually imperative that such requirements be completed if the desired services will ever be offered.

8. A Summary of the QoS Functional Areas

The management of QoS will need to provide functionality to the application and/or at the access, at the core, and at the boundaries to administrative regions.

QoS traffic functions will need to include admission control, authentication and authorization, and billing. Verification that traffic is within agreed parameters and programmatic interfaces to advise when the service is outside the agreed limits. Interfaces that provide service verification, fault notification, and re-instantiation and termination will also be necessary.

Core functions will include traffic engineering, network device configuration, fault detection, and recovery. Network devices will need to inform the management system of their available resources and the management system will need to tell devices how and where to forward data.

Between administrative regions accounting, service signaling, and service verification will be needed. At the administrative boundaries of the network functions similar to those provided at the edge will be necessary. Peer entities in different administrative domains would signal their needs across the boundary. Verification at the boundary could then occur consistent with the verification at the edge. Actual traffic through the boundaries could be measured and billing information be transferred between the domains. The central management function would be responsible for re-routing traffic in the event of a failure or to better utilize the existing network resources.

Billing requirements suggest the need for 3rd party verification and validation functions available to each provider of QoS service within the flow. On one side of the transaction functionality is needed to approve pricing and payment and on the other side there will need to be an interface to provide the pricing information and make payment request for payment demands.

These requirements will raise a host of issues not the least of which is security. For the most part security considerations will be addressed both by securing the protocols (like with IPsec) and by establishing a dedicated network for control information [6]. While it will be in most instances too costly to create a physically separated DCN it will be possible to create a virtually separated network that will provide the same security benefits. Future work in the IRTF Service Management Research Group intends to look in detail at these requirements.

9. Security Considerations

For an issue as complex as a Service Management architecture, which interacts with protocols from other standards bodies as well as from the IETF, it seems necessary to keep in mind the overall picture while, at the same time, breaking out specific parts of the problem to be standardized in particular working groups. Thus, a requirement that the overall Service Management architecture address security concerns does not necessarily mean that the security mechanisms will be developed in the IETF.

This document does not propose any new protocols, and therefore does not involve any security considerations in that sense. However, throughout this document consideration of the security issues raised by the architectural discussions are addressed.

10. Summary

The paradigm for service management in IP networks has been adopted from that of telecommunications networks. Basic differences between the service models of these networks call into question if this is realistic. Further analysis is needed to determine what is the proper paradigm for IP service management and to define a common vocabulary for it.

The IP community is currently very active in solving problems relating to transport QoS issues. These activities are illustrated by the work of the Diffserv, Intserv, and Policy working groups. In contrast not enough effort is being focused on service issues relating to applications. The present solution is for applications to build in their own service management functionality. This is

often an inefficient use of network resources, but more importantly will not provide for access to transport level services and the functionality that they offer.

The IP community needs to focus on adding service functionality that is flexible enough to be molded to specific application needs, yet will have access to service information that will be necessary to provide superior application functionality. Principal needs to be addressed relate to developing transport level services for billing and security. Directory services and extending the work done to define AAA services are promising starting points for developing this needed functionality.

11. References

[1] L. Mathy, C. Edwards, and D. Hutchison, "The Internet: A Global Telecommunications Solution?", IEEE Network, July/August 2000.

[2] B. Leiner, et. al., "A Brief History of the Internet version 3.31", revised 4 Aug 2000.

[3] Eder, M. and S. Nag, "Service Management Architectures Issues and Review", RFC 3052, January 2001.

[4] Y. Bernet, "The Complementary Roles of RSVP and Differentiated Services in the Full-Service QoS Network", IEEE Communications Magazine, February 2000.

[5] Floyd, S. and L. Daigle, "IAB Architectural and Policy Considerations for Open Pluggable Edge Services", RFC 3238, January 2002.

[6] Recommendation M.3010 "Principles for a telecommunications management network", ITU-T, February 2000.

[7] Recommendation M.3100 "Generic network information model", ITU-T, July 1995.

[8] Moore, B., Ellesson, E., Strassner, J. and A. Westerinen, "Policy Core Information Model -- Version 1 Specification", RFC 3060, February 2001.

[9] V. Jacobson, "Differentiated Services for the Internet", Internet2 Joint Applications/Engineering QoS Workshop.

[10] Nichols, K., Jacobson, V. and L. Zhang, "A Two-bit Differentiated Services Architecture for the Internet", RFC 2638, July 1999.

[11] Mankin, A., Baker, F., Braden, B., Bradner, S., O'Dell, M., Romanow, A., Weinrib, A. and L. Zhang, "Resource ReSerVation Protocol (RSVP) Version 1 Applicability Statement Some Guidelines on Deployment", RFC 2208, September 1997.

12. Authors' Addresses

Michael Eder
Nokia Research Center
5 Wayside Road
Burlington, MA 01803, USA

Phone: +1-781-993-3636
Fax: +1-781-993-1907
EMail: Michael.eder@nokia.com

Sid Nag
PO Box 104
Holmdel, NJ 07733, USA

Phone: +1-732-687-1762
EMail: thinker@monmouth.com

Hemant Chaskar
Nokia Research Center
5 Wayside Road
Burlington, MA 01803, USA

Phone: +1-781-993-3785
Fax: +1-781-993-1907
EMail: hemant.chaskar@nokia.com

13. Full Copyright Statement

Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.