

EMGT 835 FIELD PROJECT

Quality of Service (QoS) in Enterprise WAN Networks

By

Shafeeq Shaikh

Master of Science

The University of Kansas

Fall Semester, 2008

An EMGT field project report submitted to the Engineering Management Program and the faculty of the Graduate School of the University of Kansas in partial fulfillment of the requirements for the degree of Master of Science.

Name	Date
Committee Chair Herb Tuttle	

Name	Date
Committee Member Annette Tetmeyer	

Name	Date
Committee Member Linda Miller	

Table of Contents

List of Figures	iii
List of Tables	iii
Acknowledgements	iv
Executive Summary	v
Acronyms	vi
1 Purpose	1
2 Introduction.....	1
2.1 Importance of QoS for Enterprise Networks.....	2
3 Literature Review	3
4 Background	7
4.1 Overview of QoS	7
4.1.1 Reasons for QoS.....	8
4.1.2 Functions Performed by Quality of Service	9
4.2 Types of QoS Models.....	9
4.2.1 IntServ Model.....	9
4.2.2 DiffServ	10
4.3 QoS Techniques	14
4.3.1 Congestion Management.....	14
4.3.2 Congestion Avoidance	15
4.4 Types of Traffic flows	15
4.5 Strategic Approach to Design and Deployment.....	16
4.5.1 Process for Deploying QoS-Enabled Network Services.....	17
4.6 Managing the Project	19
4.6.1 Project Approach.....	19
4.6.2 Project Deliverables	20
4.6.3 Project Roles & Responsibilities.....	21
4.6.4 Work break down Structure.....	22
4.6.5 Budget Allocation	22
4.6.6 Project Management Procedures.....	22
4.6.7 Deliverable Management	22
4.6.8 Scope Change Management	23
4.6.9 Schedule Management	23
4.6.10 Quality Management	24
4.6.11 Issue Management.....	24
4.6.12 Risk Management Plan	24
4.6.13 Communications Management Plan	25
4.6.14 Project Closure Plan	26
5 Results	26
5.1 Cost Benefit Analysis	26
5.2 Summary.....	30
5.3 Conclusion	32
6 Recommendations for Further Research.....	32
7 References	33

List of Figures

Figure 1	Source: Yee-ting Lee, at the University College of London 2003.....	20
Figure 2	Source: Yee-ting Lee, at the University College of London 2003.....	22
Figure 3	Source: Cisco Systems documentation 2005.....	23
Figure 4	Source: Szigeti, Tim, End-to-End QoS Network Design 2005.....	28
Figure 5	Source: Cisco Systems, Enterprise QoS Network Design 2005.....	32

List of Tables

Table 1: Components of End-to-End Delay	8
Table 2: Industry Defined Types of Traffic	16
Table 3: High Level Project Tasks	20
Table 4: Cost Analysis of increasing WAN bandwidth due to increasing traffic requirements	28
Table 5: Cost Analysis of implementing QoS	29

Acknowledgements

I would first like to thank God for helping and providing me with the means to complete this field project report and getting to an end on my MS as a whole. After that, I thank my wife and son for showing patience while I worked on this field project report and on this MS program. I also thank my Committee chair and members for guiding me through the process and for providing support while I was writing this report in a manner that could not go without being praised.

Thank you all!

Executive Summary

Enterprise IT organizations started to face issues related to managing applications over the slow speed Wide Area Network (WAN) circuits a few years ago with the appearance and use of file-sharing applications over their WANs and/or Internet between headquarters and branch offices. The pain has continued to grow with introduction of mission critical transactional type data along with voice and video traffic over these networks.

For this very reason several techniques were developed and implemented over the years by the vendors and service providers to help the Enterprise organization cope with these problems.

This field project report provides an overview of Quality of Service (QoS) and its implementation in Enterprise Networks for Engineering Managers. The author will discuss the best practices around developing designs for architectures already in place, as well as the business and technical challenges that are faced by Engineering Managers in enterprise organizations when deploying QoS. While the project focuses on QoS implementation of WAN, similar concepts can be used for any type of implementation throughout the network.

The author will first address the issue of bandwidth utilization and how its use has increased in recent years, what type of applications are driving it and the issues enterprise organizations are having in managing it. Next, the various industry standard solutions available to tackle this issue and the advantages of deploying them in the network will be discussed. Recommendations on the use of Project Planning guidelines to implement this critical project within the company and strategic steps that can be used to accomplish the goal will follow.

Finally, cost analysis will show that increasing the bandwidth on the WAN can simply add cost to the enterprise organizations in most cases, whereas the proper use of QoS can help the organization reduce cost while utilizing the existing circuits while delivering the same level of service.

The author recommends the use of QoS in networks to manage traffic but does not rule out the possibility of increasing bandwidth in the enterprise network or the combination of both.

Acronyms

CAPEX	Capital expense
LAN	Local Area Networks
MPLS	Multi-protocol Label Switching
MRC	Monthly recurring charges
OPEX	Operational expense
QoS	Quality of Service
RFC	Request for Comments (A Standard body)
VPN	Virtual Private Network
WAN	Wide Area Networks

1 Purpose

The use of Wide Area Networks (WAN) and the Internet in Enterprise Organizations to access applications and data storage at data centers has grown rapidly. This is mostly due to organizations moving towards a centralized architecture where data is stored at the data center versus a distributed architecture in which application servers and data are locally stored at the branch office. This increase in the use of such networks has also increased the need to understand how to best utilize bandwidth for different types of applications that organization are using. A lack of understanding of the capabilities and tools that are available for such purposes have led enterprises to invest in bandwidth as their need grows. One of such tool or capability that can be enabled in network devices or applications is called Quality of Service (QoS).

With the help of QoS, Engineering Managers can manage multiple high priority, latency sensitive and best effort application traffic over the same network bandwidth. The author proposed the use of QoS as the most effective solution in an enterprise environment that brings a number of advantages to both enterprises and the service providers.

This field project report is written with the intention that information provided here would help Engineering Managers understand the basic concepts of QoS from a technology perspective and also understand the business side related to strategy, cost and project planning. Enterprises are having problems in determining if they need a bandwidth upgrade on their WAN or they would just need QoS in their network. Both options have drawbacks. Bandwidth upgrade directly increases costs while QoS requires skills to design, implement and support. This document helps Engineering Managers determine the best way to go based on their specific needs.

2 Introduction

The computing environment, in general, deals with the issues related to constrained resources. For computers, operating systems must find a way to equitably manage applications and assign resources. On network, routers and switches at the exit and entry points of the network provide such services. A few decades back, circuits or

network connections like leased lines, modem lines or ISDN were only used to transmit data from one location to another location. Similarly, there were dedicated phone lines on a separate network which provided voice services to organizations. In the last fifteen years as network speeds increased and networks becoming more reliable, the concept of converged networks has emerged. That is, voice, video and data traverse over the same network.

This converged network environment has created a challenge for Engineering Managers and Network Administrators on how to manage bandwidth requirements for each type of traffic, i.e. voice, video and data, as each have different network requirements and behaviors. Data traffic is typically tolerant of delays. For example if browsing, one would not notice a small delay on pulling the web page or if noticed is not an issue. On the other hand voice traffic has different characteristics. Voice is intolerant of jitter (delay) and packet loss. Jitter can be defined as the difference in the end-to-end delay between the packets. For example, if one packet requires 100ms to traverse the network from point A to point B and the following packet requires 125ms then the delay variation or the jitter is 25ms (Tim Szigeti, Cisco Press Book, 2005). Video is also intolerant of jitter and packet loss.

2.1 Importance of QoS for Enterprise Networks

Networks are built to generate revenues in the service provider environment and to support delivery of services and products in the Enterprise organizations. Any specific application or applications that an Enterprise or business requires to be competitive in the industry would need some sort of priority over the rest of network traffic. By managing the use of bandwidth, Network Managers can deploy QoS to make sure that above mentioned business goals and objectives are well covered.

In today's environment networks must provide predictable, measurable, and guaranteed services to the end-user application or to the business owner. QoS technologies refer to the set of tools and techniques to manage network resources and are considered the key enabling technology for network convergence (*Cisco Systems, 2005*).

The objective of QoS is to make voice, video and data convergence appear transparent not only from deployment perspective but also from end-user and application perspective. QoS would allow different types of traffic to co-exist and share the same network resources. Voice, video, and critical data applications may be granted priority or preferential services from network devices so that the quality of these strategic applications does not degrade to the point of being unusable.

For this reason, QoS is considered as an important element of the network if needed and critical to the successful business.

3 Literature Review

A Literature Review has been completed using several books and information available through the Internet or in book format. Several references used in this paper are provided at the end of this report. The most useful information was available through Cisco Press Books and at Cisco's website. Below are the citations from the books and internet articles that the author used.

According to the referenced material, there are two types of QoS models, IntServ model and DiffServ. With IntServ, resources are first reserved throughout the network as requested by the source or the initiator to its destination on the network. Once the network bandwidth/resources are reserved, the actual traffic flows takes place. The protocol that was developed to support this model was RSVP or Resource Reservation Protocol, defined under RFC 2205. RSVP enables Integrated Services, in a guaranteed, like a dedicated virtual circuit and in the Controlled Load, equivalent to "best effort service under unloaded conditions." Hence, it is "better than best-effort," but cannot provide the strictly bounded service that Guaranteed service promises [5].

Differentiated Services or DiffServ provides a simple and coarse method of classifying services of various applications. Although others are possible, there are currently two standards per hop behaviors (PHBs) defined that effectively represent two service levels (traffic classes): *Expedited Forwarding* (EF) Has a single *code-point* (DiffServ value). EF minimizes delay and jitter and provides the highest level of aggregate quality

of service. Any traffic that exceeds the traffic profile (which is defined by local policy) is discarded. *Assured Forwarding* (AF) has four classes and three drop-precedence's within each class (so a total of twelve *code-points*). Excess AF traffic is not delivered with as high probability as the traffic "within profile," which means it may be demoted but not necessarily dropped [5].

After reviewing multiple papers, research articles or vendor specific documentation, it looks like RSVP is not used a lot in the Enterprise environment but more in Service Provider environment especially with MPLS Traffic Engineering. Enterprises are leaning more towards the implementation of the DiffServ model of QoS.

As part of DiffServ model, less than best-effort, Scavenger class of traffic and a strategy for using this class has been presented in Tim Szigeti's book End to End QoS Design. Szigeti states the following:

Traffic flows can be monitored at the campus edge routers/switches and out of profile traffic can be marked down to scavenger class (DSCP CS1). To complement this policer, queues providing a less than best-effort scavenger service during the periods of congestion are deployed in the LAN, WAN and VPN [6].

The Cisco press book *LAN/WAN QoS Design* mentions the use of a strategy to deploy QoS that was not found in any other referenced materials.

The best practices strategy includes the following:

- *Clearly defining the organization's business objectives to be addressed by QoS*
- *Selecting an appropriate number of service classes to meet these business objectives*
- *Soliciting executive endorsement.*
- *Performing QoS functions in hardware instead of software whenever possible.*

- *Classifying traffic as close to the source as administratively possible or feasible, preferably at layer 3 with standard based DSCP markings.*

Successful QoS deployments begin by clearly defining organizational QoS objectives and then selecting an appropriate number of service classes to meet these objectives [6].

Some common data was uncovered regarding the typical requirements for each type of traffic. This is standard and can be found in any Cisco documentation related to the subject. As per Cisco Technical Documentation:

The service level needs of voice, video, data and the control plane are, Voice requires 150 ms one-way, end-to-end (mouth-to-ear) delay, 30 ms of one-way jitter and no more than 1 % packet loss. Voice should receive strict priority servicing, and the amount of priority bandwidth assigned for it should take into account the VoIP codec, the packetization rate, IP/UDP/RTP headers (compressed or not) and Layer 2 overhead. Additionally, provisioning QoS for IP telephony requires that a minimal amount of guaranteed bandwidth be allocated to Call-Signaling traffic [7].

Practical analysis has also shown that if the end-to-end network delay of 150ms just could not be achieved, 200ms mark can be used. In this case voice quality would be degraded.

Cisco further explains video and data traffic:

Video comes in two flavors: Interactive Video and Streaming Video. Interactive Video has the same service level requirements as VoIP because a voice call is embedded within the video stream. Streaming Video has much laxer requirements, because of the high amount of buffering that has been built into the applications. Control plane requirements, such as provisioning moderate bandwidth guarantees for IP Routing and Network Management protocols, should not be overlooked

Data comes in a variety of forms, but can generally be classified into four main classes: Best Effort (the default class), Bulk (non-interactive, background flows), Transactional/Interactive (interactive, foreground flows) and Mission-Critical. Mission-Critical Data applications are locally-defined, meaning that each organization must determine the select few Transactional Data applications that contribute the most significantly to their overall business objectives [7].

Finally, one must consider the topic of congestion management as presented in Cisco website literature:

Congestion management features allows controlling congestion by determining the order in which packets are sent out an interface based on priorities assigned to those packets. Since different types of traffic shares a data path through the network, can interact with one another in ways that affect their application performance. If network is designed to support different traffic types that share a single data path between routers, congestion management techniques should be considered [10].

Congestion management entails the creation of queues, assignment of packets to those queues based on the classification of the packet, and scheduling of the packets in a queue for transmission.

Similarly, congestion avoidance techniques should be used to avoid any type of network traffic congestion before it occurs. The difference between the avoidance and management is the latter is more related to once the congestion has occurred and former is more avoidance. The author provides detailed information on Congestion avoidance techniques avoid congestion at common network bottlenecks.

“Congestion avoidance is achieved through packet dropping. Among the more commonly used congestion avoidance mechanisms is Random Early Detection (RED), which is optimum for high-speed transit networks. Cisco IOS QoS includes an implementation of RED that, when configured, controls when the router drops packets. If you do not configure Weighted Random Early Detection

(WRED), the router uses the cruder default packet drop mechanism called tail drop. WRED is used to drop packets randomly based on the configured parameters that controls the drop probabilities from a TCP session in the network to optimize bandwidth utilization of the TCP traffic. The global synchronization problem of unguarded TCP sessions is discussed in several books along with how WRED can be used to combat this behavior [11].

It is very important for Network Engineers to understand the QoS need of an organization before making a choice between congestion avoidance and congestion management. Choosing a right tool is the key in implementing QoS in the network. Incorrect selection can cause more problems in the application or traffic flow that organization is trying to resolve.

Cisco developed WRED to add a weighted component to RED which is an industry standard technique. WRED has played major role in the industry and most widely use tool when it comes to congestion avoidance.

4 Background

4.1 Overview of QoS

Quality of service is the ability to prioritize different applications or application specific traffic types to guarantee a certain level of performance. Quality of service is important if the network capacity is insufficient, especially for real-time streaming multimedia traffic such as voice over IP and video over IP, as these types of traffic are delay sensitive. In the case that the network is over-engineered and network congestion does not occur, QoS mechanisms are not required.

The primary goal of QoS is to provide dedicated bandwidth for high priority traffic, along with controlled jitter and latency for some real-time and interactive traffic. When designing a network that is QoS aware, it is also important is make sure that when prioritizing application traffic, other lower priority applications do not fail or suffer performance degradation.

4.1.1 Reasons for QoS

The main reason for QoS would be a shortage of bandwidth due to network links that are oversubscribed or where packets are being lost due to congestion at bursty periods. These reasons could highly impact latency sensitive traffic like voice, video and even data traffic.

When determining the requirements of QoS, it is important to understand the components involved in the end-to-end delay in the network. Rhys Haden has mentioned these components in detail and explanations are provided in Table 1. "End-to-End delay made up by a number of components that occur in the following order:" [6]

No.	Component	Explanation
1	Fixed Switch Delay	As packets are layer 2 switched from the initiator
2	Fixed Encoding Delay	Coding PCM to G.729 (5 ms look ahead then 10ms per frame, with G.729 a Frame Relay packet would typically have 2 x 10ms samples), G.726(0.125us), G.723 (30ms), G.728 (2.5ms) etc
3	Fixed Packetization Delay -	i.e. packetizing samples in G.711, G.726, and G.729 etc.
4	Variable Output Queuing Delay	As voice packets enter an output queue and wait for the preceding frame (voice or data) to be played out.
5	Fixed Serialization Delay	The time it takes to get the bits on to the circuit. The faster the circuit and the smaller the packet, the less this delay.
6	Fixed Processing Delay -	The time it takes for the packets to be examined, routed, managed etc
7	Variable Network Queuing Delay	Due to clocking frames in and out of the different network switches.
8	Fixed Network Propagation Delay	Normally assumed to be 6 microseconds/km or 10 microseconds/mile (G.114) as the packets traverse the medium. Across the UK this is typically between 1ms and 3ms.

Table 1: Components of End-to-End Delay

(Haden, Rhys 1996 "Components of End-to-End delay" Available via <http://www.rhysaden.com/qos.htm>, Quality of Service. Accessed Oct 2008)

4.1.2 Functions Performed by Quality of Service

QoS serves several functions when properly implemented. It can change the way the network behaves to congestion at the WAN edge. Some of the high-level functions of QoS would be to enable [8]:

- Congestion management when it occurs
- Delay and jitter sensitive applications can have improved response times
- Dedicated bandwidth on a per application basis
- The avoidance of congestion and setting of traffic priorities
- Assign queues to each application

Now that some baseline understanding has been established, some deep diving in the topic is required to better understand the tools available. When talking about the networks there could be lot of vendors and vendor specific implementations. The standards are the same - it's just the implementation models that are sometimes different and vary from vendor to vendor. This paper focuses on the implementation of QoS by Cisco Systems on their Routers and Switches.

4.2 Types of QoS Models

There are two distinct models, IntServ and DiffServ, to help resolve quality of service issues in the network and each take care of specific needs.

4.2.1 IntServ Model

This is the model where resources are first reserved throughout the network as requested by the initiator to its destination on the network. Once the network bandwidth is reserved and resources are assigned, the actual traffic flow takes place. RSVP or Resource Reservation Protocol was developed to support this model, also defined under RFC 2205. RSVP is not covered in this document because it is not widely used. Most Enterprises use the DiffServ model which is discussed next.

4.2.2 DiffServ

The next QoS model that has been widely used, deployed and discussed in this paper in detail is Differentiated Services or DiffServ. It provides a simple and coarse method of classifying services of various applications.

In DiffServ architecture, end-to-end QoS is deployed using per-hop behavior. The differentiated service approach to providing quality of service in networks employs a small, well-defined set of building blocks from which a variety of aggregate behaviors may be built. A small bit-pattern in each packet, in the IPv4 ToS octet or the IPv6 traffic class octet, is used to mark a packet to receive a particular forwarding treatment, or per-hop behavior, at each network node. A common understanding about the use and interpretation of this bit-pattern is required for inter-domain use, multi-vendor interoperability, and consistent reasoning about expected aggregate behaviors in a network.

There are currently two standards per hop behaviors (PHBs) defined that effectively represent two service levels (traffic classes):

- **Expedited Forwarding (EF)**: Has a single *code-point* (DiffServ value). EF minimizes delay and jitter and provides the highest level of aggregate quality of service. Any traffic that exceeds the traffic profile (which is defined by local policy) is discarded. Typical use is in Voice applications.
- **Assured Forwarding (AF)**: Has four classes and three drop-precedence within each class. Excess AF traffic is not delivered with as high probability as the traffic "within profile," which means it may be demoted but not necessarily dropped.

As illustrated in Figure 1, PHBs are applied to the traffic when it enters a network according to pre-determined policy that was configured on a router or a switch. The traffic may be marked at this point, and routed, then unmarked at the network exit point. There was a development done at the end-host application, i.e. end hosts can also apply the DiffServ marking, like Cisco IP Phones and Windows operating system. There are a number of advantages in doing so but not discussed here as it's out of the scope of this paper.

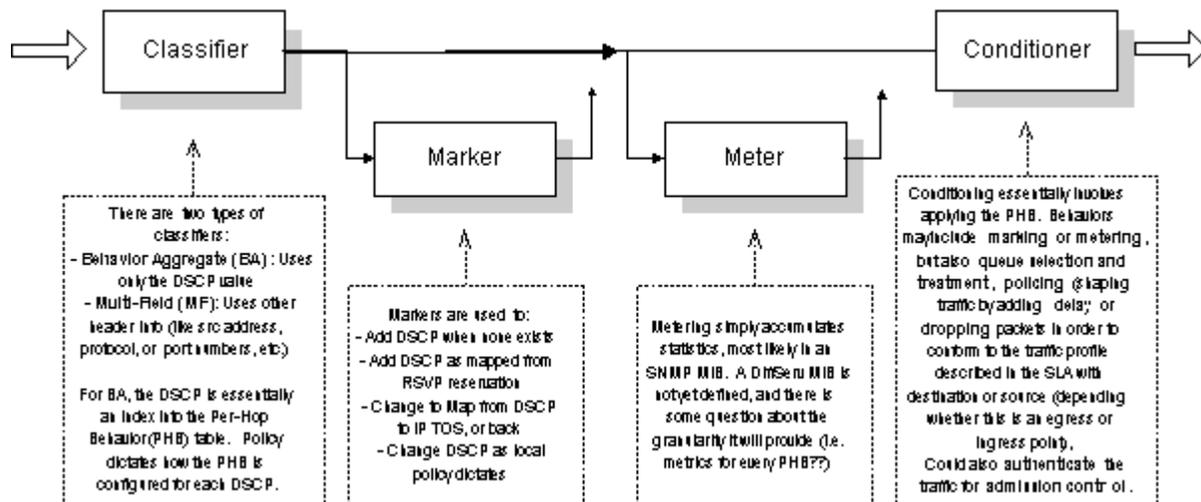


Figure 1: Differentiated Services Architecture

(Source: Yee-ting Lee, PhD paper 2003 at the University College of London, accessed Aug 2008)

Figure 1 above shows the differentiated Services Architecture, with a break out of some specifics. This functionality is enabled in every DiffServ enabled router, although not all functions are used all the time. Typically, border routers--at ingress and egress points--apply functions, but interior routers may also. As mentioned earlier, this paper covers how the DiffServ model works and shows deployment mechanisms.

Classifying and Marking

When a packet enters a network, DiffServ is used to classify it. Packets form a flow of traffic which is defined by five elements: source IP address, destination IP, source port, destination port and the transport protocol. This classification is applied to the flow of traffic.

Once the traffic has been classified, it is marked as per the defined policy.

Type of Service (TOS) is a field in the IP header that sits between the Header Length field and the Total Length field. IP Precedence (IPP) uses the first three bits of the ToS field to give 8 possible precedence values. (Refer to the IP Datagram TOS field for a view of the Type of Service field in the IP header as shown in Figure 2.) The following bit and class information is a well known standard in the work of QoS in Networking.

- 000 – Routine 0
- 001 – Priority 1
- 010 – Immediate 2
- 011 – Flash 3
- 100 - Flash Override 4
- 101 – Critical 5
- 110 – Internetwork Control 6
- 111 - Network Control 7

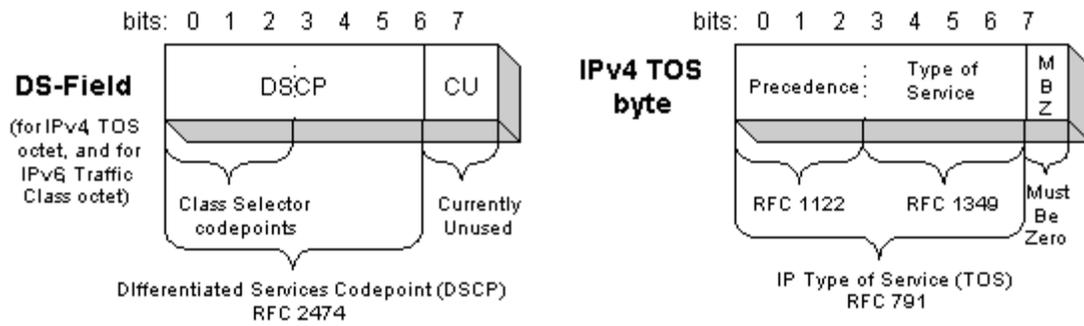
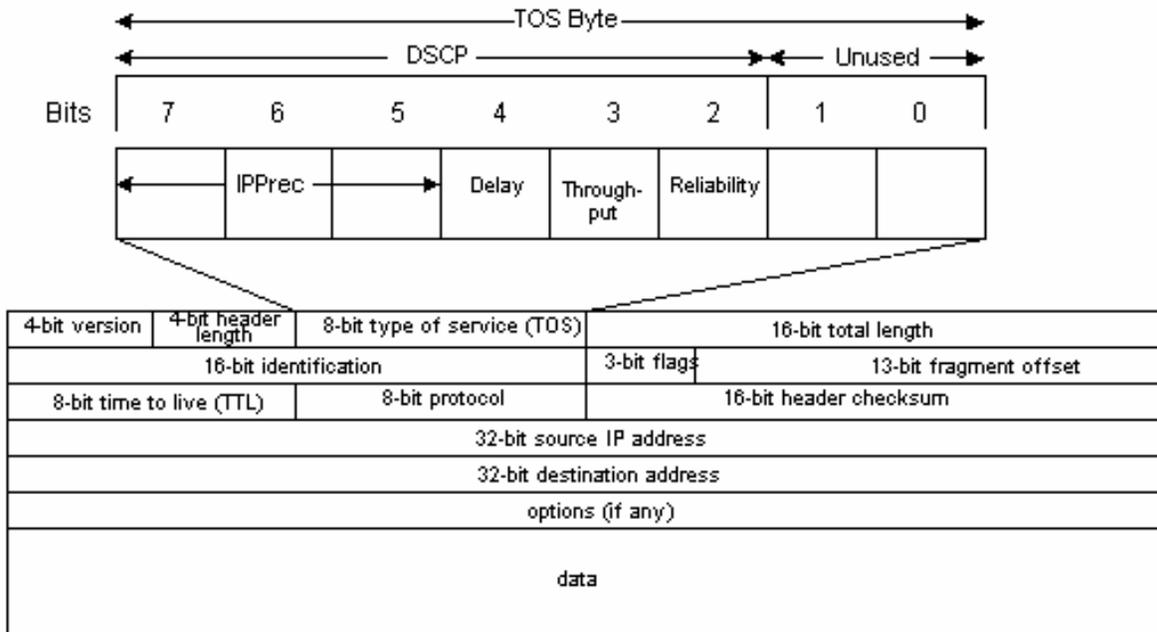


Figure 2: Differentiated Services Code Points (DSCP) redefine the IPv4 Type of Service byte. IP Precedence bits are preserved in class selector code-points & PHBs, but TOS values are not.

(Figure Source: Yee-ting Lee, PhD paper 2003 at the University College of London, accessed Aug 2008)

DiffServ also defines the concept of the DSCP or DiffServ Code Point that uses the first 6 bits of the ToS field, giving $2^6 = 64$ different values. It's important when considering IPP and DSCP markings, because they co-exist in the IP header. RFC 2474 describes the Differentiated Services (DS) field and the DiffServ Code Point (DSCP).



(Figure 3: Type of Service (TOS) byte details as part of IP header.

(Figure Source: Cisco Systems documentation, for Enterprise Organizations, 2005, www.cisco.com, pg 12)

In DiffServ architecture, each router handles each packet differently and individually, and acts upon it based on its DSCP value. That's where the concept of Per-Hop Forwarding Behavior (PHB) comes in, where classes are developed such as Executive, Business, Finance, Accounting, Teleworker, and Residential. Each class can be offered by an ISP as different levels of service or deployed within the enterprise for different types of users. The PHB has three elements to it which determine whether the router interface will:

1. Drop the datagram,
2. Send the datagram or
3. Re-classify the datagram.

This three-color marker is detailed in RFC 2697. "For example, 5 flows can be treated as a Behavior Aggregate so they are treated similarly as a group in most respects. Each flow is then distinguished by an additional Drop Probability and Forwarding Behavior. As the Drop Preference value increases, so the probability of being dropped increases" [8].

Following are AF DSCP values:

AF11 = 001010 = 10
AF12 = 001100 = 12
AF13 = 001110 = 14
AF21 = 010010 = 18
AF22 = 010100 = 20
AF23 = 010110 = 22
AF31 = 011010 = 26
AF32 = 011100 = 28
AF33 = 011110 = 30
AF41 = 100010 = 34
AF42 = 100100 = 36
AF43 = 100110 = 38

RFC 2597 defines a group of DSCP settings called “Assured Forwarding” Per Hop Behavior (PHB). The Assured Forwarding PHB class is presented as AF (xy), where x=traffic class, and y=drop precedence. 4 traffic classes and 3 drop precedence are defined. For example, AF21 = traffic class 2, drop precedence 1. The traffic class values (1-4) have escalating priority values where traffic marked as AF11 has a lower priority than AF41. Conversely, the drop precedence value (1-3) represents an escalating drop preference within the specified class, a descending priority. For example, traffic marked as AF43 is more likely to be dropped than AF41.

4.3 QoS Techniques

QoS has several techniques that were developed over the time to provide priority to certain applications. Congestion management and congestion avoidance are related to DiffServ from a high level perspective.

4.3.1 Congestion Management

Congestion management is a reactive approach, where routers and switches are configured or programmed to send packets out of the interface with defined priorities after the congestion has occurred. Queuing is done when congestion occurs at the interface of the router or a switch. It provides a way of sorting the traffic out or the packets would get discarded. Packets after being marked or colored, can be placed in queues.

There are multiple queuing mechanisms that were developed, including:

- WFQ (Weighted Fair Queuing)

- PQ (Priority queuing)
- CQ (Custom Queuing)
- Class-Based Weighted Fair Queuing (CBWFQ)
- Low Latency Queuing (LLQ)

The queuing systems are not discussed here but are mentioned to show that they are available as part of congestion management.

4.3.2 Congestion Avoidance

Congestion avoidance is a more pro-active approach where routers and switches are programmed to monitor output interface and send packets based on defined priorities and drop packets with low priority to avoid congestion on the network. Unlike congestion management, the congestion avoidance technique makes sure that congestion does not occur at first place. This is by dropping packets with low priority. Some well know congestion avoidance techniques are:

- Random Early Detection (RED)
- Weighted Random Early Detection (WRED)
- Traffic Shaping
- Traffic Policing

Again, these congestion avoidance techniques are not discussed here but are mentioned to show that they are available.

4.4 Types of Traffic flows

Table 2 shows industry defined types of traffic. Each of these traffic flows has different requirements, depending on the type and size of the network. Figure 4 shows different part of the networks each with its own limitation and requirement.

Voice	Video	Data
Bearer And Call signaling	Interactive video (Video conferencing) Streaming video (news channels, training videos)	Best-Effort, Bulk Data (FTP, Email, backups), Transactional Data/Interactive Data (SAP, PeopleSoft, and Oracle) Locally-Defined Mission-Critical Data

Table 2: Industry Defined Types of Traffic

Table 2: Industry Defined Types of Traffic

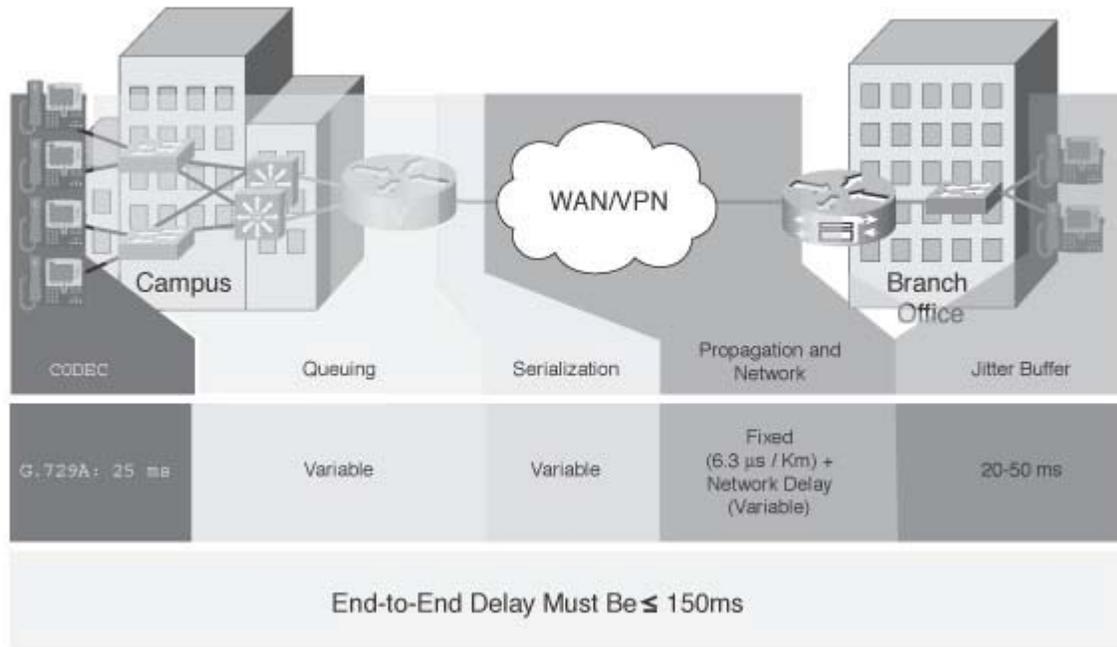


Figure 4: End-to-End Network delay.

(Source: Szigeti, Tim, *End-to-End QoS Network Design, Quality of Service in LANs, WANs and VPNs*, Indianapolis:

Cisco Press, 2005, pg 35).

4.5 Strategic Approach to Design and Deployment

Now that a basic understanding of QoS standards has been developed, discussion for having a strategic approach towards its design and deployment can be initiated. Enterprise Organizations should approach the QoS deployment as strategic and not something that can be done overnight. Strategy should be developed on how to tackle the design and deployment of the project. This will be followed by an operational support model that may increase the cost of supportability of the solution if proper skills are not already available within the team.

“The managerial process of crafting and executing a strategy consists of five interrelated and integrated phases”. [14] This strategy can be defined specifically for this project, i.e. deployment of QoS in the Enterprise organization:

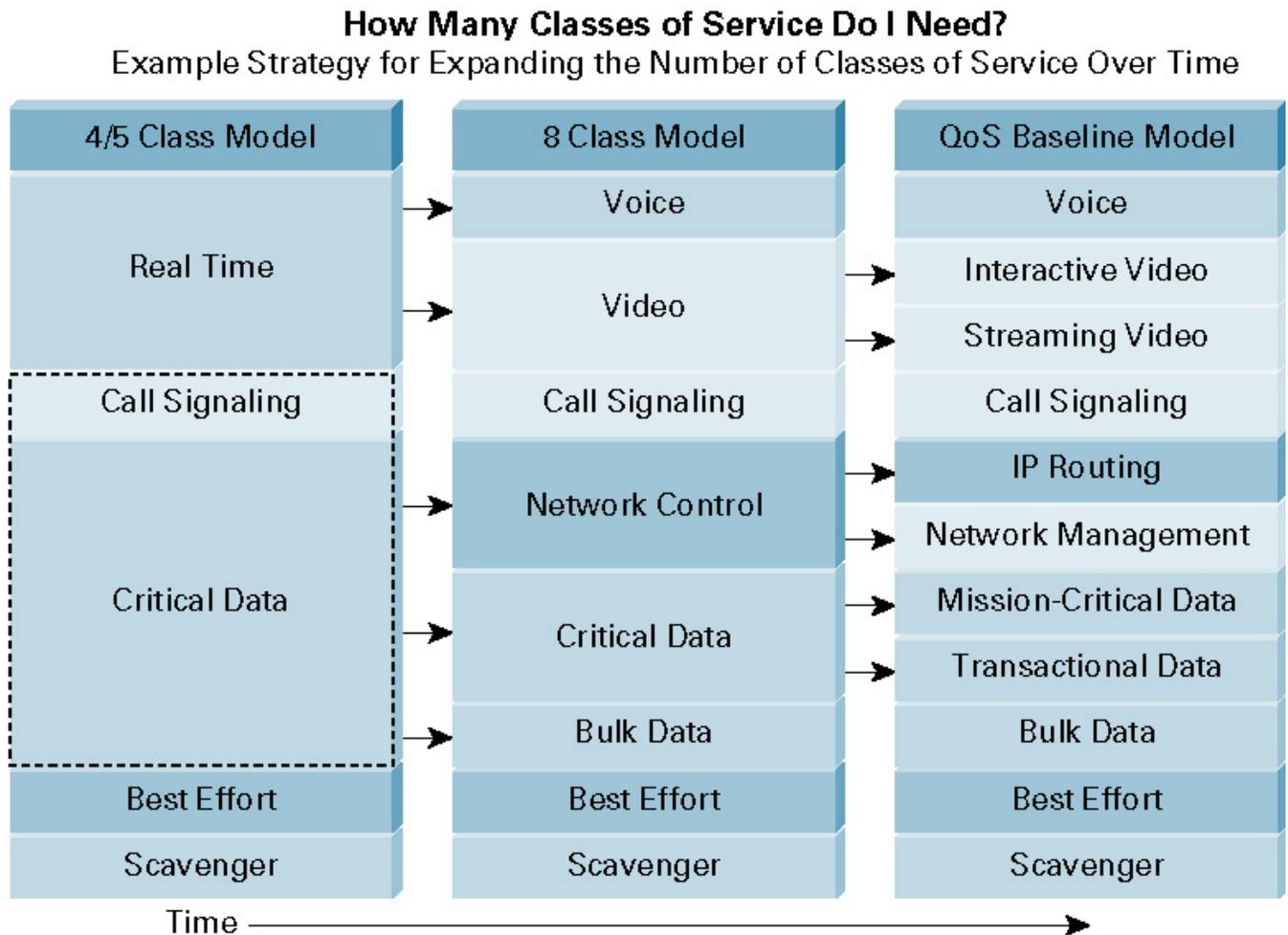
1. Set and develop a strategic vision of where the company needs to head in terms of deploying QoS.
2. Set objectives—Engineering Manager action is required to convert the strategic vision into specific goals and objectives to deploy QoS.
3. Crafting a strategy to achieve the objectives and move the company toward where it wants to go when deploying QoS.. The tests of a winning strategy are how well QoS deployment is matched to the company's requirements, whether it's producing a sustainable competitive advantage, and whether it is boosting company performance.
4. Implementing and executing the chosen QoS strategy efficiently and effectively. Managing the implementation and execution of strategy is an operations-oriented, make-things-happen activity. Translating an organization's QoS strategy into actions and results tests an Engineering manager's/Network Architect's ability to direct organizational change tied to good strategy execution and the achievement of target outcomes.
5. Evaluating performance and initiating corrective adjustments during the deployment of the project and after its deployment, shows organization's operational excellence. This phase of the strategy management process is the trigger point for deciding whether to continue or change the QoS strategy, objectives, and/or strategy execution methods. Sometimes simply fine-tuning the strategic plan and continuing with efforts to improve strategy execution suffices. At other times, major overhauls are required.

4.5.1 Process for Deploying QoS-Enabled Network Services

Companies can follow a five-step process to deploy QoS. This can be tied to the above mentioned five inter-related strategic steps:

- Define business objectives:
 - Identify mission-critical applications; Make sure that very few applications get this classification, so better end-to-end performance across the network is observed.

- Obtain endorsement on business requirements from executives.
- Determine how many classes of traffic are required to meet objectives: more classes translates to more granular service guarantees (see Figure 5 below)



(Figure 5: Strategy for QoS deployment with Time.)

Source: Cisco Systems, Enterprise QoS Design Guide document, Nov 2005 ver 3.3, pg 30)

- Analyze service-level requirements:
 - Voice, which is affected by loss, delay, and delay variation.
 - Video, also affected by loss, delay, and delay variation

- Data, which varies according to the application. Different versions of the same application can have different traffic types and characteristics. Therefore, in the case of increased data traffic or applications, QoS requirements must be re-evaluated to accommodate required service levels.
- Design and test QoS policies against business objectives and requirements.
- Roll out the QoS designs to production network in phases, during scheduled downtime.
- Monitor service levels to help ensure that objectives are met across the entire network. This step should be done in parallel with step 4 for validity of the design.

4.6 Managing the Project

To develop and implement QoS project, it's necessary to develop a project management plan for deploying QoS in the Enterprise. This section has provided some basic understanding of the Project Management policies and procedures and is an indirect interpretation of what's defined in the PMBOK (Project Management Body of Knowledge) by Project Management Institute.

It is good practice to monitor and track the project deployment of this scale, to keep tasks and money in line with each other.

4.6.1 Project Approach

The project should follow a typical system development lifecycle as defined in the organization's project schedule template. The overall project can be divided into the following high-level tasks:

Project Planning	Project Execution	Project Close
Estimation Requirements & Design Milestone: Planning Complete	Test Case Development Configuration development Functional Testing System Testing Milestone: Development Complete Deliver to QA for integrated testing (Support Integrated testing) Deploy design to Production Milestone: Execution Complete	Conduct Lessons Learned Close project Milestone: Project Complete

Table 3: High Level Project Tasks

4.6.2 Project Deliverables

Organizations should define a clear project deliverable document to set the expectations of the solution owners and the management. There is a high possibility that the vendor is involved in the project either due to the out-sourcing or because of their role in the organizational network. In this case, both deliverables should be defined.

Project deliverables also include those documents which are fundamental to the success of the project and are required to be reviewed and approved by key stakeholders. The documents listed in this section specify the project approving authority.

Product Specification

The Product Specification provides the detailed specific requirements and design for the product to be deployed. This would be determined once the survey is completed and the design is developed. It is the document that forms the basis for the scope of the project and the standard against which a change is triggered.

Test Plan/Test Cases

This document represents the acceptance criteria to demonstrate that the modifications made to the application function according to the requirements defined in the product

specification. The signatures of the project team constitutes final acceptance of the deliverable:

Test Results Summary.

As approved test cases are executed, the results of these tests are documented. This may include the status of the test steps (success or failure) and specific results of calculations, export files, screen shots, etc. Review and acceptance of the results constitutes approval to deploy the product to production:

Deployment Plan

In order to ensure that changes are implemented correctly, it is necessary to have a written plan of action. The deployment plan will specify when the change will occur, how the change will be implemented, the length of service disruption (if any), and communications required.

Test and Acceptance Plan

The test and acceptance plan is used to hand over the completed project to the organization by the vendor.

4.6.3 Project Roles & Responsibilities

The purpose of this section of the document is to describe the roles and responsibilities of the functional groups (or individuals) that are required to deliver the work packages and create the deliverable material.

Roles and responsibilities are organized into three categories: the Executive Project Team, the Core Project Team and the Extended Project Team. While the detailed responsibilities are described for each component below, the overall responsibilities can be summarized as follows:

Executive Project Team

Provides overall project direction to ensure the project is meeting the business objectives as well as ultimate decision-making authority for issue or conflict resolution.

Core Project Team

Groups (or individuals) who have direct accountability for project planning, project execution and delivery of project results to meet the defined business objectives.

Extended Project Team

Groups or individuals who have direct interest in the project and its success but are not directly accountable. Typically this group will provide consulting or resource skills that are required but do not have the level of accountability in the project as the core project team.

4.6.4 Work break down Structure

This section describes the work activities, schedule, resources, and budget details for the project. At a minimum, include a list of the key elements in the project work breakdown structure and a description of those activities. If the work breakdown is developed in elements other than “work activities,” adapt the descriptions here to conform to those elements.

4.6.5 Budget Allocation

Proper budget allocation and enough funds are necessary for the successful completion of the project.

4.6.6 Project Management Procedures

This section describes a typical Project Management Procedure that any organization would employ. The purpose here is to give a high-level of information on the subject to stress its importance as part of Project Management Plan that would be developed by an Engineering Manager or Project Manager to deploy QoS.

4.6.7 Deliverable Management

The project manager should record each deliverable. The status of each deliverable should be updated for each status reporting cycle (see the communication section for more detail).

Work on the deliverable should be assigned through the project schedule and time tracked on individual's timesheet(s).

The project manager can distribute the document for approval via Outlook email to the approving authorities for the deliverable defined in the Project Deliverables defined above. The email could contain "voting buttons" whereby the approver(s) may click either Approve or Reject to signify their acceptance or rejection of the document. Additionally the approver(s) may include comments about the document as well as signatures of authorized project team members.

4.6.8 Scope Change Management

Changes to the scope of the QoS deployment can well happen due to the criticality of the project. There may be an application that was identified initially as non-critical that now needs additional bandwidth.

Change control within the project is a key to achieving success and it is a major responsibility of the project manager. Change Management is an important control function throughout the project lifecycle. Scope change may include (but is not limited to) anything that affects the product requirements, product specifications, time, cost, resources, and/or quality.

4.6.9 Schedule Management

Project schedule control includes the tools and techniques required to monitor and control the project schedule and to ensure that the schedule status is known and understood and that the appropriate actions are taken to document and react to schedule variances. Schedule control is primarily concerned with activities associated with keeping project tasks on schedule.

4.6.10 Quality Management

Quality management is not only to determine that the product meets its functional requirements through classical quality assurance testing but also that regulatory and company-specific standards are met. This may include the following:

- All requirements must be measured against deliverables
- Prior to release for approval all deliverables must be reviewed by the core team for accuracy.
- All deliverable must be reviewed and have the appropriate signatures as identified in the Project Deliverables section, above.
- Any user interface or report changes must adhere to company specific regulations
- Any changes to external interfaces must consider the impact to SOX.

4.6.11 Issue Management

An issue is a problem situation that currently exists on the project for which a resolution must be found. Every project encounters issues that suddenly arise that must be dealt with. Every team member can raise issues and in fact the quicker they are revealed the better. For this project an issue is a known situation that if left unresolved will negatively impact the success of the project. An issue is not an action item (unless the action item is past due), nor is it a risk.

This project must utilize organizational standard issue management process and format.

4.6.12 Risk Management Plan

Risk Management is critical to the success of any project. Managing risk is not the exclusive domain of the project manager rather it is a project responsibility that is owned by each group or individual having an interest in the project.

The risk management plan is used to track events that have been predicted during the project risk assessment that can have an affect on the project plan. Typically the project team attempts to predict risk events that may adversely affect project plans and

schedules. While not all risk can be identified the intent of the risk assessment is to realistically predict those events that may be reasonably expected during the execution of the project plan. Pre and Post installation Risks are associated with this project.

Risk management issues could include managing customer downtime or long duration outages. Customer locations can have a downtime during the installation which can get prolonged if there is an issue with the configuration applied on the router and switches to support QoS. Engineering Managers should work with the vendor/Site engineers to prepare methods and procedures to avoid risks

It is important to understand the risk management process that will be employed during project execution in conjunction with this plan. There are many tools and techniques that can enhance the ability of the team to predict risk. These include personal knowledge or experience, lessons learned from past projects as well as tools such as decision trees and, certainly not the least important, the project plan itself.

4.6.13 Communications Management Plan

The purpose of the Communication Plan is to provide an overall framework for managing and coordinating the wide variety of communications that will directly or indirectly take place as part of this project. It addresses communicators, audiences, events, communication channels, feedback mechanisms and message timing. This framework is designed to provide the IT organization the quality of communications necessary for relevant, accurate, and consistent information at all times to maximize productive results.

The Program Manager must develop the Communication Plan, monitor and direct its use, and assist in the determination of its effectiveness. Actual delivery of communications will be through designated 'communicators' - presenting and facilitating briefing sessions, delivering communication and soliciting feedback.

4.6.14 Project Closure Plan

Most of the steps are the responsibility of the Project Manager who's managing the project through closure, and only those steps include a task illustration. This document also includes the steps for which the Resource Manager and IPM (IT Portfolio Management) are responsible, thereby providing a complete step representation of the project closure process.

5 Results

5.1 Cost Benefit Analysis

Every project has an associated cost. Management performs a cost benefit analysis to try to determine if the project is going to help the business earn more revenue and/or value or if the project will only increase cost. Based on these results, decisions are normally taken to either to move forward or stop the project. For this reason, it is important to understand both the cost of deploying QoS and potential savings that the project could generate.

Cost of bandwidth always depends on the type of service required and the amount of bandwidth needed. For example, lower speed circuits like T1 (1.5Mb) are much cheaper than DS3 (45Mb) or higher.

A simple cost analysis to determine if increasing the bandwidth v/s deploying QoS would be beneficial or not can be done by looking at each individual cost that would be incurred during all phases of the project lifecycle.

There are two types of costs that could be associated with this project, CAPEX and OPEX. CAPEX represents capital expensing that would be needed to deploy this project within the enterprise. The OPEX represents an operational cost needed to support the solution once deployed and operational along with reoccurring charges.

The author suggests finding the cost associated with current circuits utilized by an organization. Additional costs associated with upgrading current bandwidths to

accommodate traffic requirements also need to be estimated. This cost is usually available from the service provider that the organization is using for its Wide Area Network connectivity.

Let's assume that an organization ABC has 100 sites connected on a Frame-Relay MPLS/VPN network with a service provider. These 100 sites might consist of:

- Four Data Centers with OC12 (622 Mbps) type connectivity into the MPLS network.
- Fifty sites with T1 i.e. 1.5Mbps type connectivity to the provider MPLS network.
- Forty sites with 3Mb connections i.e. 2xT1
- Six sites would be assumed to be larger locations having 10Mb Frame circuits.

IT Organization within company ABC has been receiving complaints lately, about slow response times on certain mission critical applications recently launched, especially from their remote sites having T1 connectivity. The application server resides at the Data center. After a detailed analysis of the bandwidth usage report, it has been discovered that the use of internet at these sites is causing congestion at the WAN edge routers and delay in the application response. Two options came up to help resolve the problem. First, increase the WAN bandwidth at these sites to accommodate the requirement. Second, deploy QoS. The QoS option brings another long term opportunity, i.e. it would build the foundation for deploying Voice over IP (VoIP) solution in a future.

In order to make a decision, cost analysis is required to determine the cost of deploying QoS and the cost of upgrading the circuits. It is highly recommended that QoS should be used when deploying VoIP or Video over IP type solutions in the network.

In this example, author first determines the cost of increasing the bandwidth at the sites other than the data center, because the data centers already have high bandwidths and assumes that the research was done, which shows that data centers do not need upgrades at this time.

Design

Let's assume that the design is suggesting all the sites with T1 type connectivity would need doubling of the bandwidth to support the voice and video traffic with a limited numbers of calls along with the mission critical applications. Sites with 3 Mb type connections would need 4.5Mb connections and sites with 10Mb connections would need to be upgraded to 15Mb connections.

Cost of Increasing Bandwidths/QoS

Table 4 shows that adding bandwidth to the current circuits would increase monthly recurring costs by almost \$90,000. This does not include labor cost or hardware cost that might be needed in some upgrades and/or build-out cost. Even if the organization decides to upgrade the communication circuits and is ready to pay these costs, it does not guarantee the timely delivery of voice, video and data traffic according to their priorities or need. Engineering managers must determine how to justify this cost or recover it by giving applications more bandwidth. Research has shown that in most cases more bandwidth means more internet based non-business traffic which could negatively impact the performance of the applications that are critical to the business.

Number of Circuits	MPLS/VPN Frame Relay Circuit Type	Current Cost	After upgrade to 3MB	After upgrade to 4.5MB	After upgrade to 15MB	Current Total cost # of Circuits X Current Cost	Total Cost after upgrade # of Circuit X Upgrade Cost
50	T1	\$2,090	\$3,040			\$104,500.000	\$152,000.000
40	2xT1	\$3,040		\$3,990		\$121,600.000	\$159,600.000
6	10M DS3	\$5,996			\$6,955	\$35,976.000	\$41,730.000
MRC						\$262,076.000	\$353,330.000

MRC = Monthly Recurring Cost

Table 4: Cost Analysis of increasing WAN bandwidth due to increasing traffic requirements

Engineering Managers can now determine if deploying QoS in the organization WAN circuits would serve the purpose of delivering business applications within scope, schedule and cost.

Quantity	MPLS/VPN Frame Relay Circuit Type	Current Cost	Circuit Cost with QoS	Current Total cost # of Circuits X Current Cost	Total Cost after QoS # of Circuits X Circuit Cost with QoS
50	T1	\$2,090	\$2,299	\$104,500.000	\$114,950.000
40	2xT1	\$3,040	\$3,344	\$121,600.000	\$133,760.000
6	10M DS3	\$5,996	\$6,595	\$35,976.000	\$39,570.000
MRC				\$262,076.000	\$288,280.000

MRC = Monthly Recurring Cost

Table 5: Cost Analysis of implementing QoS

Table 5 shows the cost analysis of deploying QoS in the network. QoS can be deployed with a slight increase in an overall cost of operating the network, but the benefits are long term. Usually, service providers charge 10% of the cost of the circuit to provide QoS based services that would guarantee the preferential treatment of packets in their network and delivered within Service Level Agreement.

It is very clear that QoS can be deployed with only a slight increase in monthly recurring charges. This increase in charges can be easily recovered with revenue generated by the timely delivery of critical application data. For example, retail stores can deploy QoS to make sure that the customer service application that they use over the WAN network would have better response times so more customers can be serviced which could generate more revenue. Similarly, organizations can deploy video applications for conferencing purposes between distant locations by implementing QoS over their WAN network. The enhanced video experience would bring cost savings by minimizing executive and staff travel costs.

Circuit charges are not an actual presentation of circuit cost or cost in total, but numbers based on industry average. A similar model can be used for calculating real cost. The

author suggests including any cost related to hardware, labor and build-outs to more accurately estimate realistic costs.

5.2 Summary

The purpose of this report is to provide the Engineering Manager with some basic knowledge of QoS and a deployment strategy that would help them make a decision based on their organization needs.

An explanation of Quality of Service and why it is needed over WAN (Wide Area Networks) was provided. This project report provides an overview of Quality of Service and its implementation in Enterprise Networks for Engineering Managers. Best practices around developing designs for architecture already in place as well as business and technical challenges that are faced by Engineering Managers in enterprise organizations when deploying QoS were discussed.

The use of internet/WAN based services is really the driving force behind the use and requirement of QoS in the Enterprise organizations. Applications are increasingly being centralized to the data center and users who were accessing applications locally are now accessing the same applications over their WAN/Internet network. This has increased the need of understanding how to manage bandwidth on the Wide Area Network connections. One option is to buy more bandwidth from the service provider to accommodate the increase in application traffic. But for most organizations, this may not be the best solution. The organization needs is to prioritize traffic in such a way that critical application traffic takes precedence over discretionary activities such as internet browsing, movie downloading or streaming video.

With the help of QoS, Engineering Managers and Network Engineers can manage the bandwidth among the high priority, latency sensitive traffic and best effort traffic. Likewise, congestion management and congestion avoidance tools can be utilized to overcome the above mentioned issues, along with classification and marking tools, policing tools and scheduling tools.

QoS could be optimally deployed within the enterprise by using a five phase approach, presented as part of strategic deployment.

The author recommends the use of project planning guidelines to implement this critical project and has provided some strategic steps that can be taken to accomplish the goal. At the end, the author employed a cost analysis to prove that increasing the bandwidth on the WAN can simply add cost to the enterprise organization whereas the proper use of QoS can help to reduce cost while utilizing the existing circuits to deliver services that are now more responsive, predictable and reliable.

The report further addresses the need of using project management tools to control and monitor the project. Project management procedures were also defined which play an important role in the overall project. These were:

- Deliverable Management
- Scope Management
- Schedule Management
- Quality management
- Issue Management
- Risk Management

The author concluded by addressing the highly-relevant question: How does cost justify the deployment and support of QoS v/s just increasing the bandwidth? This is shown using the example of an enterprise network that wants to deploy voice and video along with the data traffic over their WAN networks. The analysis clearly specifies that in most cases (not all) it is financially and technically feasible to deploy QoS in WAN networks to support multiple mission critical applications along with voice and video as business requirements.

5.3 Conclusion

The author of this field project report has tried to convey that upgrading bandwidth is not always a solution to a problem of over utilizing bandwidth and to the deployment of voice, video and data over the same network. Proper requirements are needed to determine the right course of action. The author concluded that QoS is important for Enterprise Networks that are moving towards deploying voice, video and data applications over the same, single and current network infrastructure. This is feasible related to both the business and applications.

6 Recommendations for Further Research

This project report has not covered the LAN (Local Area Network) or Campus QoS and therefore, the author recommend doing research in these areas. Usually QoS is not deployed in the LAN or Campus environment because high capacity Ethernet bandwidths are available and links do not get congested. However, QoS is highly recommended in this case in order to achieve a high quality voice and video experience. Also when QoS is deployed on the WAN, it is a good idea to implement it in the LAN environment as well. Each of the following research areas would have different cost structure and cost impact on the enterprise. Research could include Campus QoS Design, MPLS VPN QoS Design and IPSec VPN QoS Design.

7 References

Standards/Books/Internet

1. J. Postel. RFC-791: Internet Protocol. Available via <http://www.rfc-editor.org/rfc/rfc791.txt>, Sept. 1981.
2. J. Heinanen. RFC 2597: Assured Forwarding PHB group. Available via <http://www.rfc-editor.org/rfc/rfc2597.txt>, Jun 1999
3. B. Davie RFC 3246: An Expedited Forwarding PHB (Per Hop Behavior). Available via <http://www.rfc-editor.org/rfc/rfc3246.txt>, March 2002
4. Cisco Technical Documentation. "Enterprise QoS Solution Reference Network Design Guide". Available at <http://wwwkr.cisco.com/univercd/cc/td/doc/solution/esm/gossrnd.pdf>. 2005.
5. Yee-ting, Lee. 'Quality of Service' Available at <http://www.hep.ucl.ac.uk/~ytl/qos/index.html>. 2003. University College of London (accessed Aug 2008)
6. Szigeti, Tim and Hattingh, Christina. 2004 "End-to-End QoS Network Design: Quality of Service in LANs, WANs and VPNs". Cisco Press. Pg 63 (accessed Sept 2008)
7. Cisco Technical Documentation. "Quality of Service Design Overview". Available via http://www.cisco.com/en/US/docs/solutions/Enterprise/WAN_and_MAN/QoS_SRND/QoSIntro.html. Nov 2005 (accessed Oct. 2008)
8. Haden, Rhys. "Quality of Service" 1996. Available at <http://www.rhysaden.com/qos.htm>.
9. Cisco Technical Documentation. "Deploying Control Plane Policing". Available via <http://www.cisco.com/warp/public/732/Tech/security/docs/copp.pdf>. 2004

10. Cisco Technical Documentation. *"Congestion Management Overview"* Available at http://www.cisco.com/en/US/docs/ios/12_2/qos/configuration/guide/qcfconmg.html. April 2001. pg 1 (Accessed Oct 2008)
11. Cisco Technical Documentation. *"Congestion Avoidance Overview"* Available at http://www.cisco.com/en/US/docs/ios/12_2/qos/configuration/guide/qcfconav.html. April 2001 (accessed Sept 2008) pg 1-2
12. Flannagan, Michael, Richard Froom, and Kevin Turek. 2003 *"Cisco Catalyst QoS: Quality of Service in Campus Networks"*. Cisco Press
13. Sziget, Tim and Hattingh, Christina. 2004 *"End-to-End QoS Network Design: Quality of Service in LANs, WANs and VPNs"*. Cisco Press
14. Thompson Jr, Arthur A, John E. Gamble and A.J. Strickland III 2004. *"Strategy: Winning in the Marketplace: Core Concepts, Analytical Tools, Cases"* The McGraw-Hill Companies.
15. Kerzner, Harold. 2006. Project Management *"A systems approach to planning, scheduling, and controlling"* John Wiley and Sons. Hoboken, NJ
16. Ross, Stephen A, Randolph W. Westerfield and Bradford D. Jordan. 2004 *"Essential of Corporate Finance"* The McGraw-Hill Companies.