NETWORK PERFORMANCE MANAGEMENT USING APPLICATION-CENTRIC KEY PERFORMANCE INDICATORS

by

SUSAN E. MCGILL
B.S. Barry University, 1971
B.A.S. Florida Atlantic University, 1983
M.S. Illinois Institute of Technology, 2001

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Modeling and Simulation
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Summer Term
2007

Major Advisor: Randall Shumaker

ABSTRACT

The Internet and intranets are viewed as capable of supplying “Anything, Anywhere, Anytime” and e-commerce, e-government, e-community, and military C^4I are now deploying many and varied applications to serve their needs. Network management is currently centralized in operations centers. To assure customer satisfaction with the network performance they typically plan, configure and monitor the network devices to insure an excess of bandwidth, that is over-provision. If this proves uneconomical or if complex and poorly understood interactions of equipment, protocols and application traffic degrade performance creating customer dissatisfaction, another more application-centric, way of managing the network will be needed.

This research investigates a new qualitative class of network performance measures derived from the current quantitative metrics known as quality of service (QOS) parameters. The proposed class of qualitative indicators focuses on utilizing current network performance measures (QOS values) to derive abstract quality of experience (QOE) indicators by application class. These measures may provide a more user or application-centric means of assessing network performance even when some individual QOS parameters approach or exceed specified levels.

The mathematics of functional analysis suggests treating QOS performance values as a vector, and, by mapping the degradation of the application performance to a characteristic lp-norm curve, a qualitative QOE value (good/poor) can be calculated for each application class. A similar procedure could calculate a QOE node value (satisfactory/unsatisfactory) to represent the service level of the switch or router for the current mix of application traffic.

To demonstrate the utility of this approach a discrete event simulation (DES) test-bed, in the OPNET telecommunications simulation environment, was created modeling the topology and traffic of three semi-autonomous networks connected by a backbone. Scenarios, designed to degrade performance by under-provisioning links or nodes, are run to evaluate QOE for an access network. The application classes and traffic load are held constant.

Future research would include refinement of the mathematics, many additional simulations and scenarios varying other independent variables. Finally collaboration with researchers in areas as diverse as human computer interaction (HCI), software engineering, teletraffic engineering, and network management will enhance the concepts modeled.

To my unofficial committee:

Dr. Ann Miller, Dr. Kitty Eeltink, Diane Scholten, and M’Lisse Ramsey

my very accomplished and supportive friends.

Each of you inspires me and together you have been the wind beneath my wings!

ACKNOWLEDGMENTS

First and foremost I must acknowledge my great good fortune in having Dr. Shumaker as my Advisor. He and my extraordinary committee challenged me to reach beyond my comfort zone and achieve something greater than I had believed possible. While all my committee members were generous with their time and encouragement Dr. Hartman and Dr. Clarke were key enablers of my research, as well as the campus Network Operations team who were also very generous with their time and expertise

Finally a deep sense of gratitude for Dr. Li, director of the NSF IGERT program at CREOL in the College of Optics and Photonics, Dr. Rolland, who first believed in me, and to the extraordinary, mind-boggling, graduate students at CREOL.

This work has been supported in part by NSF under grant number 0114418

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS/ABBREVIATIONS

ATM	Asynchronous Transfer Mode
ANSI	American National Standards Institute
BER	Bit Error Rate
BN	Backbone Network
DES	Discrete Event Simulation
FTP	File Transmission Protocol
HCI	Human Computer Interaction
IP	Internet Protocol
IT	Information Technology
ITU	International Telecommunication Union
ITU-T	Telecommunication Standardization Sector
ITU-R	Radio Communication Standardization Sector
ITU-X	Internet Communication Standardization Sector
Gbps	Giga bits per second
HTTP	HyperText Transmission Protocol
KPI	Key Performance Indicator
LAN	Local Area Networks
MIB	Managed Information Base
NOC	Network Operation Center
POTS	Plain Old Telephone System
PSDN	Packet Switched Data Network
PSTN	Packet Switched Telephone Network
QOE	Quality of Experience
QOE_{app}	Quality of Experience per application
QOE_{node}	Quality of Experience per network node
%QOS	Percent Quality of Service
QOS	Quality of Service
SLA	Service Level Agreement
SNMPv2	Simple Network Management Protocol currently v2
TCP	Transport Control Protocol
TCP/IP	TCP over IP
UDP	User Datagram Protocol
VLAN	Virtual LAN
VPC	Virtual Permanent Circuit or Virtual Path Connection
WAN	Wide Area Network
WLAN	Wireless LAN

CHAPTER 1: OVERVIEW OF RESEARCH

We have a hunger of the mind which asks for knowledge of all around us, and the more we gain, the more our desire; the more we see, the more we are capable of seeing

Maria Mitchell: Astronomer, 1st woman inducted into the
American Academy of Arts and Sciences, 1848

In the 1990s the convergence of plain old telephone system (POTS), and data networks increased the complexity of network management. Before this convergence POTS had been optimized over decades and managed for a single application -voice transmission- initially transmitted on analog switched circuits, moving to digital packet switches. Performance management on the POTS network was primarily carried out by engineers.

After the 1970s data-related networks, i.e. local area networks (LANs) and wide area networks (WANs) initially developed for a few applications like file transfer, email, and client/server programs, proliferated. Data networks led to widespread use of digital packet-switching technology. Performance management for data networks became the responsibility of Information Technology (IT) departments.

The present day wireline, wireless, and optical networks are also converging with a proliferation of multimedia applications (voice, data, and video), in addition to the existing applications for voice and data networks. The increasing availability of the Internet and World Wide Web, make telecommunications networks increasingly central to life in a global society. Global competitiveness is pushing the telecommunications industry toward ubiquitous computing, “anything, anytime, anywhere” locally and globally which must be managed to the customer’s satisfaction.

A key unknown for network researchers and network providers is the nature of the application traffic that will be generated by high social demand for some type or types of telecommunication and consequently what resource and performance demands will be placed on the network. While unknowable in detail, this application traffic is certain to be growing in volume and complexity, demanding changes to the infrastructure that will only increase future network management challenges and make performance quality increasingly difficult to guarantee. This research investigates a function to produce a performance indicator that is application sensitive and at the same time infrastructure insensitive since this would enable network management to assure performance quality as customer satisfaction becomes ever more business critical.

Based on a network-centric view, i.e. transmission infrastructure and data traffic, there have been two major research areas to aid network management: first: quality of service (QOS) parameters or metrics. QOS parameters measure the performance of bits, bytes, packets in terms of data transmission figures of merit, e.g. Bit Error Rate, Packet Loss Ratio, etc. Second, there are transmission protocols which give explicit priority to packets insuring quality transmission for those applications. There are currently several protocols which partition and allocate network resources to guaranty performance by including Class of Service (in ATM networks) or Type of Service (in TCP/IP networks) priority information in the headers of packets.

In view of these advances and faster switching and transmission media, some network researchers take the position that bandwidth will become infinite and cheap and/or that protocols and equipment can make the networks so robust and reliable that application performance degradation will not reach a user perceivable level.[1] This simplifying assumption, useful in the near term or perhaps long term in managing the core network, is already being questioned due to indications that the TCP/IP access networks are frustrating their users in spite of being well engineered. The position taken in this dissertation is that there will always be scarce resources and unforeseen consequences from protocol and data streams interacting in an increasingly complex network infrastructure. The potential for degraded performance for some or all of the applications using the network at any given moment in time will continue (and perhaps even increase) due to emergent, i.e. unforeseen and unanticipated, behaviors degrading some types of application performance and creating customer dissatisfaction.

Quality of Experience (QOE) indicators, as opposed to strictly technical metrics could serve to trigger network management, centralized or distributed, to take timely remedial actions to maintain network performance. Consequently, network management will need a measurement, qualitative in nature, to indicate the level of application, and by extension the degree of user satisfaction, provided by the network. It is a basic tenant of any process improvement program that “what isn’t measured, can’t be managed.”

This dissertation investigates qualitative indicators of the impact the network on satisfactory application performance. Theses application-centric indicators are based on existing quantitative network performance measures. Further, the utility of these indicators as a trigger for network management performance analysis and possible corrective action is demonstrated. This approach considers the direct impact on individual user experience and may result in a more robust and flexible network management. While actual use of these indicators is outside the scope of this dissertation a few simulation scenarios of application affecting network performance problems are examined to determine the sensitivity of the indicators to network conditions.

Descriptions of three proposed application-centric performance indicators follow.

To measure the network contribution to application performance, the percent quality of service (%QOS) value is proposed. It directly relates current network QOS values to the application specified QOS requirements. This is a quantitative measure, intentionally non-dimensional in nature, consequently not expressed in units of time such as per second or per millisecond.

The various %QOS values for a single application constitute a vector indicating how well or how poorly that application's performance requirements are being met by the network. When this vector is transformed by an lp-norm like function it will be designated QOE_{app}. It indicates in a qualitative manner the over all performance, good or poor, of each application type of interest.

Finally, to serve as an indicator of the current level of satisfaction or dissatisfaction at any layer two or three network node (switch or router) there will be a QOE_{node} value. This is a qualitative measure of network provided satisfaction for the current mix of application traffic. The current traffic mix is represented by a vector of QOE_{app} values, and again transformed to give a qualitative measure of customer satisfaction with the performance of the network node.

The key point is that we seem to be moving into an environment in which networks must be managed with an application-centric, or customer-centric, view, rather than network-centric only. This is a big change. The performance indicators proposed are designed to aid network management when investigating and evolving the optimal network management policies, device configurations, or network expansion plans driven by application-centric customer needs. Thus, network management will be empowered, focused, and responsive to the needs of business customers.

Adding indicators focused on application performance to the current network-centric performance metrics will add some of the information needed to move Network Management into proactive management maturity. Some business management researchers propose process improvement models for Network Management organizations that require a greater focus on becoming a profit center rather than an overhead organization consumed by operations and maintenance of the network. Expertise and insight into customer application performance should enable a deeper understanding of user needs and their demands on the network. In this way IT departments can plan for and propose market differentiating services with the flexibility to meet whatever opportunities and demands the next generation applications present to the telecommunications infrastructure.

1.1. Background

1967 – Plain Old Telephone System (POTS) is the only telecommunications interface consumers and the vast majority of businesses, universities and governmental agencies know. "Ma Bell" is the provider and "someone, somewhere" takes care of the infrastructure. It's grown to be a commodity so no one has cause to think about what the infrastructure involves or any metrics associated with running it.

1987 – While the majority of consumers still rely on POTS for their telecommunications needs, the landscape in corporations, universities and the government has changed dramatically with the introduction of telecommunications networks. Local Area Networks are pervasive and have changed the landscape of work dramatically. But even more dramatic is the effect of Wide Area Networks, allowing institutions and the users within those institutions to communicate with colleagues all over the world. These networks, initially developed for a few applications such as File Transfer, Email and Client/Server programs proliferated widely and rapidly. Suddenly issues like Quality of Service, network load factors and performance across multiple computers, often in geographically dispersed locations (including internationally) became a day-to-day issue that impacted the work lives of anyone who worked in a white collar workplace of any size. Now it was not someone "out there" who occasionally had to "tweak" the normally invincible phone system. The new world of telecommunications was up close & personal in every worker's life and required highly trained technicians within every company. Network management became more essential and more complex. Deficiencies in managing Quality of Service for telecommunications highly impacted the business/academic/governmental environment.

2007 – During the past twenty years the telecommunications landscape has been radically transformed. From grade school children through their great-grandparents, computers, the Internet and a plethora of heretofore unimaginable multimedia device and applications are a day-to-day part of the life of many people all over the world. On the consumer side, people routinely send music, photographs as well as text from varying communication devices (computers, Personal Data Assistants, Cell phones) to single or multiple users all over the world. Multiplayer online games, the ability to search through computers all over the world to find information in seconds that used to take highly trained librarians days or weeks – everyday life has been transformed by technology. On the business/university/governmental side, the results are equally dramatic. Medical personnel routinely send X-Rays to specialists over the network. Online medical surgeries using Telecommunications networks to drive robotic "arms" on the other end are even becoming more commonplace. Large military simulations of battles, research being carried out in Universities on 3 different continents – all of these large-scale applications, communications and interactions - totally unimaginable by the average person just a decade ago, are now commonplace.

With this rich new world of communication, interaction and multimedia comes a complex new set of network management issues. With POTS no one but a few “Ma Bell” engineers somewhere worried much about Quality of Service, network management or load factors. When LANS & WANs proliferated in corporations, universities & government, a new cadre of highly trained technicians in the back rooms of those institutions scrambled to keep up with the burgeoning proliferation of applications being developed to utilize this new seemingly abundant resource of “bandwidth”. Suddenly the white collar worker was affected by network outages – if the LAN or WAN “hiccupped” critical deadlines could be missed. Workers came to understand that they relied more on these network engineers than they had previously realized. Industry responded by highly rewarding people with these skill sets.

Then came the “Internet revolution”. Suddenly computers became nearly as ubiquitous as TVs. “Regular people” relied on wide area telecommunications for connection with others, entertainment, learning, job searching. Very much of day to day life has been transformed by telecommunications directly or indirectly.

With this ubiquity came an explosion of applications to fill that same seemingly limitless bandwidth. But suddenly, with the explosion of both users and applications, bandwidth management became more of a conundrum. As distributed applications, such as multimedia of all types, massively multiplayer online games, grid computing and simulations become more central to the fabric of global society, it is imperative for network performance management to measure and monitor some new application-centric quality of experience (QOE) parameters to supplant the prior Quality of Service (QOS) metric which has been the barometer of network efficacy for the past 20 years.

The present day wireline, wireless, and optical networks are also converging with a proliferation of multimedia applications (voice, data, music and video) in addition to the existing applications for voice and data networks. The increasing availability of the Internet and World Wide Web make telecommunications networks increasingly central to life in a global society. Global competitiveness is pushing the telecommunications industry toward ubiquitous computing/communicating – “anything, anytime, anywhere” both locally and globally which must be managed to the customer’s satisfaction.

And that’s just what we now know! New uses and means of accessing telecommunications network are sure to rapidly arise. Network management needs to change as radically as the uses of the network have.

This dissertation proposes just such a radical revamping of network management philosophy, techniques and metrics. It examines in more detail the development telecommunications usage and network management historically, examines the current pressures brought to bear upon the telecommunications industry and network managers as this “information revolution” continues, and proposes a new paradigm for network management – application-centric management that focuses on the Quality of Experience (QOE) variable, rather than the less-nuanced “Quality of Service” heretofore used.

The rest of chapter 1 amplifies the research with a more complete discussion of telecommunications performance management ending with an overview of the structure of the dissertation.

1.2. **Framing the Research**

The following discussion is provided to frame this research for those not familiar with networks and their operation. It helps to establish the context within which the research question is meaningful and where this research makes a contribution.

1.2.1. Network as Socio-technical System

As data networks become more central to the functioning of society in this "Information Age," the need to deploy manageable, reliable, robust, and evolvable networks is becoming as important as effectively and efficiently moving bits from source to destination. In the first years of the 21st century, the National Research Counsel (NRC): Computer Science and Telecommunications Board Committee on IT Research in a Competitive World noted a need to shift some of the focus in Information Technology (IT) research to functions typical of network management. More research is needed that focuses on the communications infrastructure as a large socio-technical system[2], which will be discussed in the next section. This type of research is required to help avoid the massively expensive IT failures that are all too common. [3, 4]

Similarly, the NRC Committee on Research Horizons in Networking felt the need to include researchers from various disciplines that are heavy users of networked applications into the report on future network research funding, to gain a broader socio-technical perspective. They concluded that new abstractions, and perhaps simplifications, were required to develop a better understanding of network performance beyond the component models, which is where the majority of research was, and mostly still is, focused. They advocate funding and encouragement of research in several key areas: measurement, modeling and building prototypes which could be considered disruptive to the current Internet assumptions. Furthermore they advocate a shift from the data plane to the management plane.

> Over the past three decades, several bodies of theory, such as performance analysis and resource allocation/optimization, have contributed to the design and understanding of network architectures, including the Internet. However, as the Internet has evolved into a critical infrastructure used daily by hundreds of millions of users, operational concerns such as manageability, reliability, robustness, and evolvability have supplanted performance of the data forwarding plane as the limiting factors. Yet theoretical understanding of these crucial areas is poor, particularly in comparison with their importance. [5]

While this dissertation focuses on access networks, not the Internet, the refocusing of research to the operational plane is needed, perhaps even more urgently, since the access networks are closer to the customers or end-users. Those who manage access networks must balance the social demands on their network and the technologies supported in the network infrastructure. These considerations can be examined within a socio-technical framework, a technique for studying systems where organizations and technology interact. While organizational development is not the subject of this research it is the rationale and would be impacted in a major way if network management shifted from network-centric to application-centric indicators such as the ones proposed in this dissertation.

Socio-technical systems theory began in the 1950s with Eric Trist's analysis of the coal mining industry since the time-motion analytical methods of Frederick Taylor's scientific theory of management were proving inadequate. Since then socio-technical theory has evolved and informed research in systems engineering, information systems, and organizational development. Weibe Bijker of MIT in the pivotal work "*Of Bicycles, Bakelites, and Bulbs: Toward a Theory of Socio-technical Change*" formalized socio-technical theory and generated a good deal of research in this area.[6-8]

Harvard’s Clayton Christensen adapts this theory “*Innovator's Dilemma: When New Technologies Cause Great Firms to Fail* “ to the process of disruptive innovation with three interdependent stages: social group has needs, a variety of solutions are generated, and one solution dominates since it is simple and easy to use when filling the need. This becomes the generally accepted solution and is incrementally improved until a new solution (or technology) meets the need better.[9] In the network arena this would describe the series of innovations that led from the Department of Defense’s (DoD) original need for a single application, a file transfer protocol (FTP) that would give them needed command control over nuclear assets in the event of a Cold War conflict, to the TCP/IP Internet of today with its multiplicity of applications. This cycle could be generalized as: (1) an application need is identified; (2) researchers, entrepreneurs, and vendors provide solutions; (3) a *de facto* or *de jure* solution and business model becomes dominate, until that solution no longer fills the need conveniently starting the cycle all over again.

A socio-technical system model was developed to clarify the rationale for this research into a QOE indicator. The first axis is the social needs and the technology enabled solutions to meet those needs. Technological solutions are developed by computer scientists, contributing advances in the design of distributed applications and network protocols, and by physicists and engineers, contributing improved devices and transmission media.

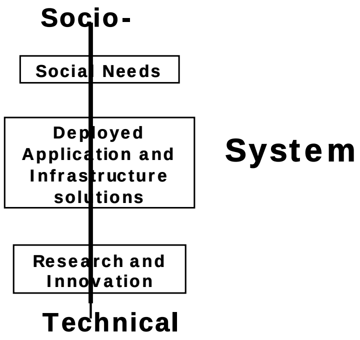


Figure 1 Socio-technical Axis

There is a chicken and egg relationship between society and technology. Social needs can be met in any number of ways, Technology provides some ways to meet those needs, and service providers and vendors create a profitable business by economically matching the demand with cost effective technology infrastructure. As more and more organizations or individuals place demands on the infrastructure the initial design has to be supplemented or enhanced until it becomes unwieldy and the next disruptive technological innovation carries the socio-technical system forward yet again. As is so often the case with technology, it's an iterative process.

The needs of organizations and individuals (expressed as applications) drive the traffic on the network leading to the question “Who are the network users and what do they want?” This business model question is the second axis of focus for this research: consumers and providers.

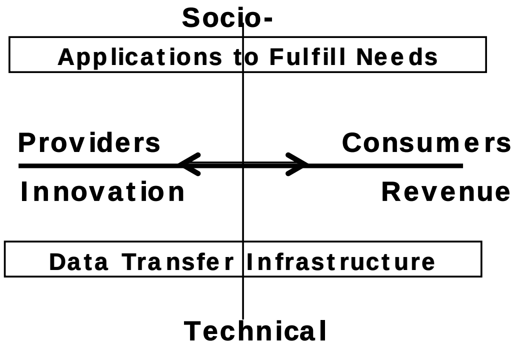


Figure 2 Business Model for Second Axis

Consumer demand drives business, and then the business providers demand that scientists and engineers supply technologies for them to use to meet these societal demands. At the same time the success of a new technology, such as the telephone or wireless communication, leads society, both the organizations and individuals, to envision meeting new needs, or they place an unanticipated magnitude of demand on the deployed infrastructure. User dissatisfaction grows until such a time as the advent of a new disruptive technology, like the computer or the all optical network, once again provides solutions resulting in a surplus of network capacity. Then as described above the multiplicity of infrastructure solutions compete until the markets stabilize on a *de facto* or perhaps a *de jure* optimal solution. Then the cycle begins all over again.[9]

At the intersection of these two axes (social needs met with technological solutions and the consumers and providers of infrastructure or services) lies the task of network management.

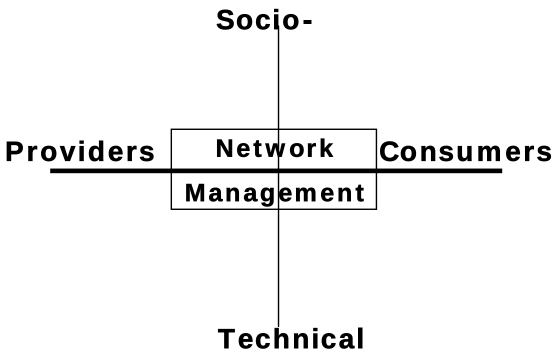


Figure 3 Task of Network Management

Although entrepreneurial business services and network technologies are changing rapidly, network mangers still have the responsibility of operating and maintaining the network to the performance level required by the customers. This level can be either implicitly, by what consumers will accept, or as the business matures, in a formal service level agreement (SLA) with specific performance failure penalties.

Initially, network management was the province of network engineers who realized in the early 1980s that metrics were needed to measure application-affecting network parameters like bit error rate (BER), delay, lost data, etc. As the networks matured network providers realized that different classes of service could be offered at different tariff rates, or price points, for applications needing performance guaranties. The Asynchronous Transfer Mode, (ATM), standard was the first network telecommunications architecture, originating in the telephone network in the mid `80s, to offer classes of service such as constant bit rate (CBR) for applications sensitive to variable delay and willing to pay a premium for guaranteed throughput real time variable bit rate (rtVBR) for applications that do not tolerate delay in general, and three other classes. At one time it was thought that ATM would be the *de facto* standard for all networks, both wide area networks (WAN), or the "core" network, and the local area networks (LANs) or access networks. Thus quality of service (QOS) would be universal through the ATM standards of quality. [10]

At about the same time the vendors of LAN equipment, originating in the information technology (IT) world of computers and based on TCP/IP over Ethernet protocols, began offering switches that were a more cost effective solution than the massive switched networks from the world of telephony[1]. Consequently the networks accessing the core, also known as edge or access networks, were left without QOS, being TCP/IP based rather than ATM based, and offered only a "best effort" delivery mode. For early applications like FTP, email, and data base transactions this was not a major impediment to quality. Although interactive distributed data base applications suffered from delay in the query/response, as do web services today, it was never clear whether the delay was a network issue or a client/server application issue. Initially, this finger pointing didn't matter as much as it does today. Since end-users have become more sophisticated and there are more choices in service providers finger pointing doesn't help if the customer is dissatisfied. From a business perspective the customer just moves their business elsewhere, creating what is referred to as "churn" in the provider's revenue stream. Business service providers have consequently brought pressure to bear on both the application suppliers and network suppliers to resolve performance issues. This has led to two major TCP/IP QOS solutions with a lot of standards activity but no clear *de facto* solution broadly adopted by most equipment vendors. On the application side it has led to a greater focus on human-centric application design and a deeper consideration of network performance on application performance.

Within the area of QOS for TCP/IP networks two major solutions are contending for acceptance at this point. Integrated Services (IntServ or IS) and Differentiated Services (DiffServ or DS). [11, 12] Both of these protocols are based on the management of bandwidth according to the class of service that the customer purchases or specifies as sufficient to avoid excessive latency, jitter, dropped packets, etc. for some anticipated mix of application traffic

The other research area has its roots in human factors, POTS telephony, and software usability. It is referred to as end-to-end user quality of experience (QOE).[13, 14] While the term "Quality of Experience" is not uniformly adopted it refers to the end-user satisfaction with the performance of networked applications as a whole. Sometimes this is referred to as the business (vs. technical infrastructure) quality of service or some other term, such as user perceived quality (UPQ). While it is clear that QOE is the overarching metric for measuring network performance, historically it has been hard to measure. Consequently the easy to measure technical metrics of data transmission (QOS) have almost exclusively been used.

Considerable human factors research was done in the POTS to determine the requirements for the development of codecs (compressor-decompressor) for analogy and digital voice applications. This resulted in the Mean Opinion Score (MOS), a subjective measure, which was used to survey end-users regarding the fidelity and quality of voice transmissions.[15] This subjective measure has evolved over the decades to a more objective programmatic test of data transmission metrics and maps the signal characteristics to the subjective scale developed in MOS. Today there are ITU standards for perceptual analysis measurement system (PAMS) and perceptual evaluation of speech quality (PESQ) commonly used to determine VOIP quality.

Many human factors studies are being performed today to determine the experience of different classes of application users. What is acceptable delay when waiting for a web page to download? What application and network factors impact a virtual reality user and produce “sim-sickness”? [16] Computer scientists are also involved in QOE research to design better algorithms, protocols, and architectures to ensure the optimal performance of distributed applications. An example of that type of research is Services Oriented Architecture which has resulted in web-based application algorithms and protocols like the service oriented architecture programming/protocol (SOAP).[17]

For enterprise IT departments and service providers research into QOE falls under the rubric of customer relationship management (CRM). This has resulted in organizational best-practices, CRM systems and specific recommended organizational skills for capacity planning, ease of order entry, ease of service deployment, i.e. provisioning or order fulfillment, and strong customer support in the form of order tracking, help-desks, and the building of knowledge bases to facilitate responses to customer enquiries. [18]

In the area of network operations and maintenance the main QOE focus is ease of device configuration management and fault management to support reliability. This has been based on network management systems developed according to standards which primarily facilitate interoperability of devices. The two standards most prevalent today are the telephony originated Common Management Information Protocol/Guidelines for Definition of Managed Objects (CMIP/GDMO) or the data communications TCP/IP based Simple Network Management Protocol/Managed Information Base (SNMP/MIB). CMIP/GDMO is one part of a more complete ITU model of network management referred to as the Telephone Managed Network (TMN) [19].

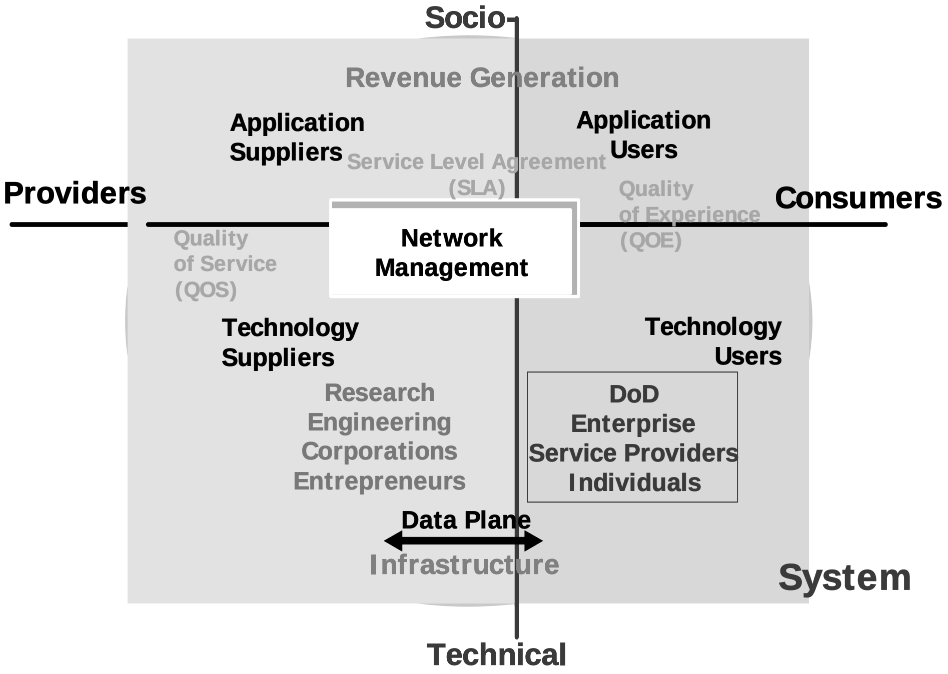


Figure 4 Focus of Network Research

Two other network management schemes exist and have been adopted into the TMN specification.[20] They are the EDI for e-commerce and CORBA for object oriented applications network management. It is widely anticipated that SNMP [21] will also be fortified to handle the complexity of global network management and be incorporated into TMN.

Figure 4 summarizes the discussion in this section with a sampling of the major players in the field and their principle concerns or contributions. To date the push for research and innovation and has been from the lower half i.e. the infrastructure, leading to advances in transmission media, protocols, and algorithms for distributed client/server or service oriented applications. It's not surprising that network management has been reliant on management protocols (SNMP and GDMO) and management statistics that are very device oriented and based on averages of aggregate traffic through nodes and links. Today the push seems to be shifting to the upper half of the diagram as content is becoming as important as infrastructure and may become the driving force for the next generation of infrastructure.[14, 22, 23]

Businesses and network management struggle to understand how device oriented MIB measurement data can be analyzed to demonstrate that the level of quality guaranteed in the customer service level agreement (SLA) is indeed being provided. Since these metrics are averages for all traffic the letter of the SLA, expressed as QOS metrics, may be complied with but the customer may be dissatisfied with how individual mission critical applications are performing. This research aims to provide QOE indicators which express QOS values in an application-centric manner.

1.2.2. Network Infrastructure and Quality of Service (QOS)

Since the QOE indicators proposed in this dissertation are based on mathematical transforms and mapping of the QOS metrics details of the network infrastructure and QOS metrics are in order. Willinger *et al* in a Colloquium of the National Academy of Science [24] when speaking on the difficulty of modeling the Internet propose a validation framework for models that are explanatory not just evocative of the emergent behavior which today challenges network researchers. First make the model data driven; second, devise mathematical constructs in terms of networking elements or mechanisms; finally close the loop by predicting behaviors that should be found in the data based on the new model. QOE indicators are for monitoring performance so scenarios have been created to close the loop and each mathematical term has been linked to network artifacts.

1.2.2.1. Infrastructure Considerations

The convergence of telephone networks (wireline and wireless) and data networks (intranets and the Internet) has been occurring since the `90s. Telephony networks were refined over decades as a single application service, voice communication. Data networks originated from large enterprise and organizational mainframe to data entry terminal and mainframe to mainframe communication, IBM BSC (Binary Synchronous Control, X3.4 circa 1965) evolving into SNA (System Network Architecture circa 1975). With the advent of Dataphone Digital Services (DDS widely deployed in the ‘70s) these data applications began transmitting over the telephone network’s media, using leased lines. Dial-up modems and Ethernet LANs began appearing in the early `80s. It was only the late `70s that Vint Cerf, Bob Kahne and Robert Metcalfe began ARPANET (the first seeds of the Internet) with TCP/IP and Ethernet. Data networks have always had varied applications, such as file transfer, distributed data bases, and email, and those application needs have spawned many variants of the basic TCP/IP over

Ethernet protocol (Reno, New Reno, ECN, etc.). The World Wide Web (WWW) began with the CERN browser in 1992 and the rest is history! [25]

The Internet backbone and interconnects are currently the main focus of much QOS network research to achieve faster, more robust and reliable data transfer. Yet access networks are much closer to the end-user and represent a getter opportunity for equipment vendors and Internet or intranet service providers. Research in QOE is of greater value in the access network and QOS may continue to be the management metrics for backbone and core networks. There are well known and standard port assignments for the most common applications like email, file transfer, web browsing, etc. Other applications use IT assigned ports. The traffic will be analyzed on a per port basis (application-centric) for interarrival times and message size. As a technical infrastructure the common abstraction of a hierarchical tree structure of access and core nodes will be used. At one level workstations will be considered access and the switch core. At the next level switches will be an access point and the campus backbone core. This abstraction will facilitate performance management using the key performance indicator proposed in this dissertation. Figure 5 shows the hierarchy of networks based on KC Claffy's dissertation[26] and highlights the research areas of interest to this dissertation.

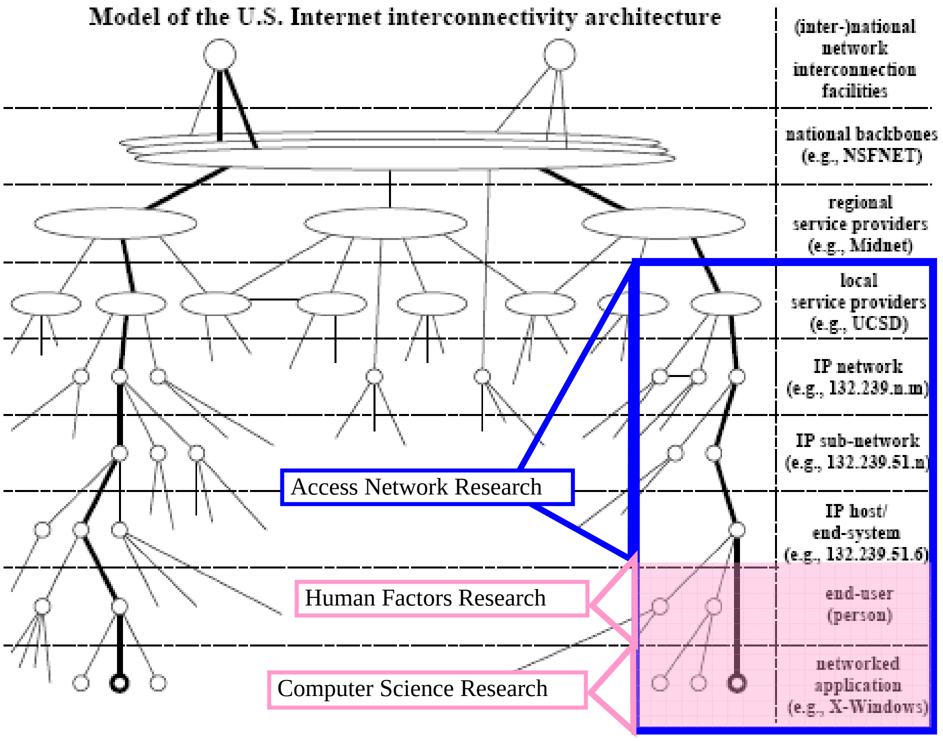


Figure 5 Hierarchy of networks highlighting areas of QOE research

The nodes (circles) represent computers or switches while the networks (ovals) contain both switches and routers. The links (lines) represent the transmission cables which are referred to as the physical layer and, based on the characteristics of the copper or optical fiber, will carry data at various speeds, expressed as bits per second (bps), commonly known as bandwidth. This dissertation deals primarily with the switches and routers that collect QOS data and store it in the local MIB or GDMO data base (Figure 6).

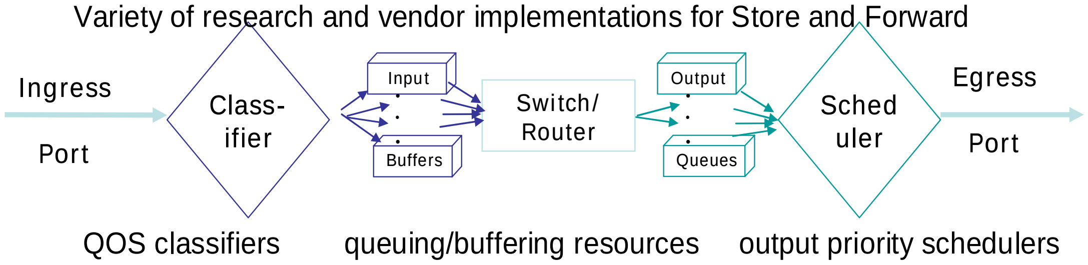


Figure 6 Node QOS Components

QOS data is collected by the network operators using SNMP or CMIP requests. This is known as active monitoring and requires bandwidth to pass messages and data back and forth. As will be discussed further in the next section, this keeps intelligence (data and decisions) in the edge devices controlled from centralized operations. Current research is in the area of distributed network management where intelligent agents collect data, perform analysis, and take corrective actions. This scheme is already being implements by some vendors with data collection and analysis being performed by "appliances" strategically located throughout the network. One example is the Cisco Network Analysis Module (NAM) card installed in some switches and routers. Cisco and researchers are also looking to predictive simulations as network analysis tools. Currently Cisco offers an off-line overnight analysis service, Network Application Performance Analysis (NAPA), which suggests corrective actions based on predictions from a simulation of current network configuration and traffic.

1.2.2.2. <u>Basic Philosophies for Providing Network Performance Management</u>

In the convergence of telephony and data networks two very distinct architectures have to be integrated and the application demands on the two systems are very distinct. Voice communication was the major application for telephony with FAX and data modems as a late development. The Internet and intranets are viewed as capable of supplying “Anything, Anywhere, Anytime” and e-commerce, e-government, e-community, and military C^4I are now deploying applications to serve their many and varied needs. Wireless is yet another architecture participating in this telecommunications convergence and has evolved since the 90s to service all types of applications.

Fred Baker, Cisco Fellow and former chair of the Internet Engineering Task Force (IETF) in the foreword to [27], characterizes POTS as an “Intelligent Network” pushing complexity and intelligence into the network while leaving the edge devices simple, i.e. telephone hand sets. This creates a very user friendly situation but places the burden of adding functionality on the network provider. Enhancing functionality means adding new devices or software systems, effectively replacing the network, at considerable expense. To recuperate those expenses various classes of service and tariffs, or price premiums, are established. Under this type of tight control quality is engineered into the network. As an example of the impact of this philosophy Baker cites the move in the mid-60s from analog to digital voice, which also allowed limited data applications (FAX) and digital video over digitally conditioned leased lines. Digital trunk lines, T1 and the whole digital hierarchy of tariffed bandwidths, and Switching System 7, a digital switch, replaced the core analog network as far as the local central office. Analog was left for the edge or “last mile” with dialup modems eventually creating low bandwidth connectivity.

As the usage grew it was once again torn out in the 1990s and bandwidth enhancements achieved through implementations of new core Asynchronous Transfer Mode (ATM) switches for wirelines, Synchronous Optical Network (SONET) and Fiber Distributed Data Interface (FDDI) for optical links and more digitally conditioned lines, T1 etc. To offer these services to businesses a distribution layer between the core and the edge was added. ATM was the first attempt at guaranteeing transmission quality with different classes of service, at different price points providing an attractive business proposition that made the investment worthwhile.

High bandwidth to the home is the current infrastructure challenge being met by the telephone companies with DSL services to compete with cable networks DOCSOS data services. The last upgrade to the core network resulted in a glut of optical fiber links and bandwidth, so that filling the available core bandwidth, or as referred to in the industry “lighting the dark fiber”, drives the business planning of some companies. Today bandwidth increases are expected to be met by tearing out the core and deploying photonic switching and Dense Wavelength Division Multiplexing (DWDM) for an “All Optical” transport layer. Currently the wireless data networks, WiFi and others, are being deployed as a less expensive technical solution and a more lucrative business model.

On the other hand, The Internet (and all TCP/IP intranets) are based on an End-to-End principle, as characterized by Baker, where intelligence is at the end, or edge, nodes with network interface cards (NIC), transmission protocol stacks and sophisticated algorithms in the applications to deal with the “best effort” service provided by the network. This architecture was initially designed for interoperability of heterogeneous devices and operating systems and the two main applications envisioned at the time were; transfer of large files and replacement of CRT terminals (hard wired to the computer) with virtual terminals which could “telnet” to multiple computers. The power of this design triggered the fulfillment of social needs with applications such as email, user groups, messaging, and distributed data base applications which were adequately served by the “Best Effort” quality of service with guaranteed delivery of data.

In the mid 1990s the WWW and web browsers, based on TCP/IP over Ethernet, became wide spread and triggered an unforeseen amount of innovative uses. Among the most demanding video and voice applications, i.e. streaming media, video conferencing, interactive multimedia, networked simulations and games, and VOIP. These applications all have much more stringent requirements for end-to-end quality of service beyond what the end nodes can control. The End-to-End design philosophy of TCP/IP has been very successful in allowing data communications between heterogeneous, standards based devices and applications but this success has resulted in such high social demand for sophisticated applications that network researchers and technology providers are forced to provide more intelligence in the network.

This success has led to a telecommunications network which is highly engineered, well understood in it's components, large scale, and a complex system that has continuously surprised researchers with emergent behaviors, a characteristic of chaotic systems. Willinger et al observe, "it is still surprising how often networking researchers observe 'emergent phenomena'- measurement driven discoveries that come as a complete surprise, can not be explained or predicted within the framework of the traditionally considered mathematical models…". As the social demands on the system shift from device performance to focus on the performance of complex applications network managers will be faced with more emergent phenomena impacting their quality. This dissertation examines composite application-centric indicators to serve as triggers to deeper analysis.

Today's network-centric measures of performance (see Table 1: Network-Centric Quality of Service Metrics), are expressed in terms of aggregate bits, bytes, or packets from undifferentiated application traffic, and do not directly reflect the performance of any application or class of applications as experienced by the customer.

Table 1: Network-Centric Quality of Service Metrics

Key QOS Network Performance Metrics (average value per measurement period)	
Delay: various statistics	EtoE: End to End (includes all delay) Propagation: through physical layer Transmission: bandwidth limited Service Time: at each node Queuing Delay: may include input & output queues depending on equipment
Delay Variation low variance allows algorithms to compensate	Jitter: variability of delay
Lost Packets some protocols retransmit	Packet Loss Ratio: # lost / total packets
Bit Errors some protocols retransmit or error correction algorithms may compensate	Bit Error Rate (BER) # bits flipped / total bits in period
Throughput: various statistics	Packets Sent Bytes Sent Bits Sent

ATM, Integrated Services, and Differentiated Services (see Appendix A for details) move toward greater customer satisfaction by establishing network-centric CoS (Class of Service) and ToS (Type of Service). This allows the application owner to select from the existing service offerings the QOS metrics and priority that will be given to their application traffic and enter into a SLA (Service Level Agreement) based on average network performance.

1.3. Overview of Subsequent Chapters

Chapter 2 reviews in greater detail: Network Management operationally and organizationally; the role of simulation in network research; Quality of Experience Research. Chapter 3 lays out the methodology for both the QOE indicators and the simulation test-bed.

Chapter 4 reports on the simulation results from both the test-bed validity perspective and the utility of the proposed QOE indicators. The finally chapter discusses the contribution of the proposed indicators and future work.

CHAPTER 2: REVIEW OF RELATED TECHNOLOGIES

"If it's not measured it can't be managed."
6 Sigma slogan (anon)

Entirely new classes of applications with supporting protocols, hardware, and transport media are proliferating and none of the old applications are going away.. The current explosion of voice, video, web services, and other new uses for the communications network has shifted the focus of organizations and vendors from the deployment of critical infrastructure to the generation of revenues from new applications made possible by the existing infrastructure. Next Generation wireline and wireless architectures, protocols, and equipment are focused on different classes of service which will have distinct billing rates. [22, 23, 28] While this migration is taking place network operations must still manage the deployed network to meet the Service Level Agreements already in place with its customer base or risk losing business. In the case of an enterprise network, failure to meet internal customer expectations means a failed project which may have been critical to the success of the enterprise (government or corporate). [29] The following sections highlight some of the active areas of research that impact the manageability and evolvability of networks which is the main concern of this dissertation.

2.1. Network Management

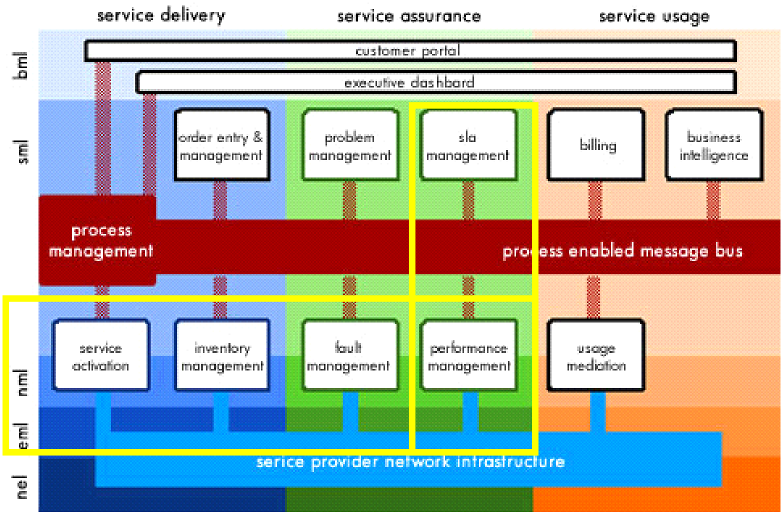
This is a broad area that will be divided into operations functional responsibilities, centralized vs. decentralized, and the organizational skills and role of the network management. All are areas under considerable pressure and active research.

2.1.1. Functional Responsibilities

To help define the issues Figure 7 depicts the whole area of network management and outlines in yellow boxes the area of performance management under consideration. The network infrastructure is managed in a hierarchically decoupled fashion as defined by the International Telecommunication Union – Telecommunications Network Management – Telecommunications Management Network(ITU-TMN) management standard[20]. The major tasks of network management are on the top axis, commonly known as provisioning, operations and maintenance, and billing. The infrastructure layers, on the y axis are, from the bottom:

- nel - network element layer (network equipment and transmission media)
- eml - element manager layer (where some future Intelligent Agents might reside)
- nml - network management layer
- sml - service management layer
- bml - business management layer

The interior of the figure shows the product architecture viewed by one vendor, Hewlett Packard, of network management products for large enterprises. This research focuses on the eml and nml for the KPI research.



http://www.openview.hp.com/solutions/serviceproviders/index.html

Figure 7 A Centralized Network Management Platform

The IETF standard SNMPv3 protocol is used to communicate to the network elements, or devices, and collect data from the Management Information Base (MIB). Standard definitions for data in the MIB specify a number of QOS metrics that should be available. The vertical yellow box captures the interrelationship between network performance management and service level agreements with users of the networks. This is the explicit or implied quality the user applications can expect from the network.

Today networking research is shifting to system level enhancements for manageability without losing its focus on enhancing hardware or protocol components for increased reliability and a robust data plane. There is evidence of this shift in the National Research Council's (NRC) report in 2001 [5] calling for networking research into modeling performance at a higher level of abstraction and using simulation on large (>100k nodes) networks two difficult and underappreciated areas of research. Additionally recent workshops in disciples related to Wireline, Wireless and Optical networks [30-33] there are references to the need for enhanced performance management. Industry is also moving in this direction with new performance management products and features [22, 23, 28]. To quote from a tutorial on Performance Management at the International Engineering Consortium Web ProForum site

> "…an increase in demand for services with new performance characteristics, demands that service providers go beyond the traditional fault management approach to performance management. Service providers must now manage their network resources to optimize the performance of the services they deliver – a fundamental change that demands an integrated performance management solution." [34]

Integrated performance management is currently added to the infrastructure as distributed monitoring and analysis appliances from vendors such as Cisco Systems [35] and partnerships between companies like Intelliden and IBM and HP network management platforms. The addition of autonomous intelligent agents is in research [36, 37] and will soon reach vendor products. Shift from Centralized to Decentralized Network Management Network management is currently centralized (see Figure 8 Centralized Network Management) with more or less automated alarm handling (Fault Management).

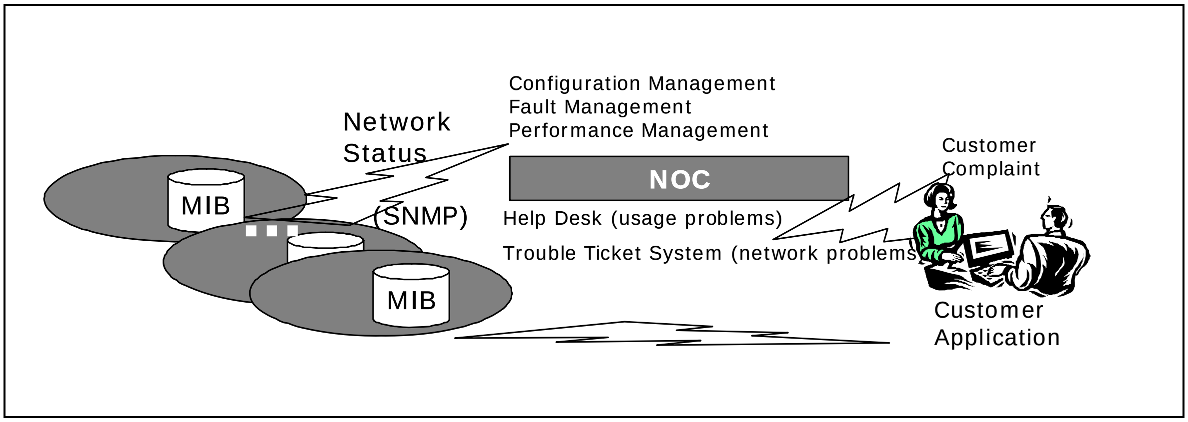


Figure 8 Centralized Network Management

Another major function of network operations and maintenance is to optimize the network performance autonomously. This may include techniques such as traffic shaping to avoid congestion, and detecting and troubleshooting security breaches. As vendors enhance the device functionality for advanced queuing or security features the task of configuration management becomes more challenging and time consuming.

The main source of network performance metrics is from the MIB (Managed Information Base) of each device using the SNMP (Simple Network Management Protocol. The SNMP / MIB standard specifies collection of per port traffic statistics but they are rarely used for monitoring network health in a centralized management scheme. For the NOC to collect this level of detailed statistics on a regular basis from the devices would create additional traffic reducing bandwidth available to application traffic. It becomes more practical if the distributed network management schemes being proposed are indeed implemented. Fred Baker, of Cisco Systems and a member of the Internet Advisory Board, explains in the foreword to [27] when the relative merits of three network management protocols were debated by the IETF (which eventually adopted SNMP/RMON) he was struck with the fact that what mattered was not so much the protocol used to communicate device status to network management but the intelligence used to analyze and remediate performance problems which is the crux of the matter. Some form of artificial intelligence commonly rule based reasoning, decision analysis, and fuzzy neural nets either currently implemented or soon to be deployed in various fault and security management applications with configuration management being the current focus of attention.

The research trend toward distributed network management with Intelligent Agents either at the devices, as a monitoring appliance, or traveling the network monitoring for performance anomalies leads to interest within the research, the standards bodies and vendors in New Generation Operations SubSystems (NGOSS) such as policy based network management. Typically these architectures consider standardized network objects and rule based policy languages for agent to agent and agent to centralized manager communications. Monitoring and evaluating device or subnet performance for well understood failure modes taking re-configuration remedial steps. Several have suggested incorporating simulation as a form of predictive feedback control. A simulation could predict the expected value of the performance metrics given past traffic history and current configuration. Detection of an anomaly would trigger rules for analysis and corrective actions.[38]

2.1.2. Organizational Role of Network Management in Quality of Service

Significant shifts are taking place in Network Management. Organizationally it has been considered a cost center or overhead of doing business. This created a good deal of focus on managing the devices in the network, through configuration and fault management, so that they performed as advertised. This resulted in ad hoc or reactive, fire-fighting, network management which simply reacted to customer complaints. In a survey of the current (2004) state of network management Gartner Group maps the organizations performance to a 5 stage capabilities maturity model.

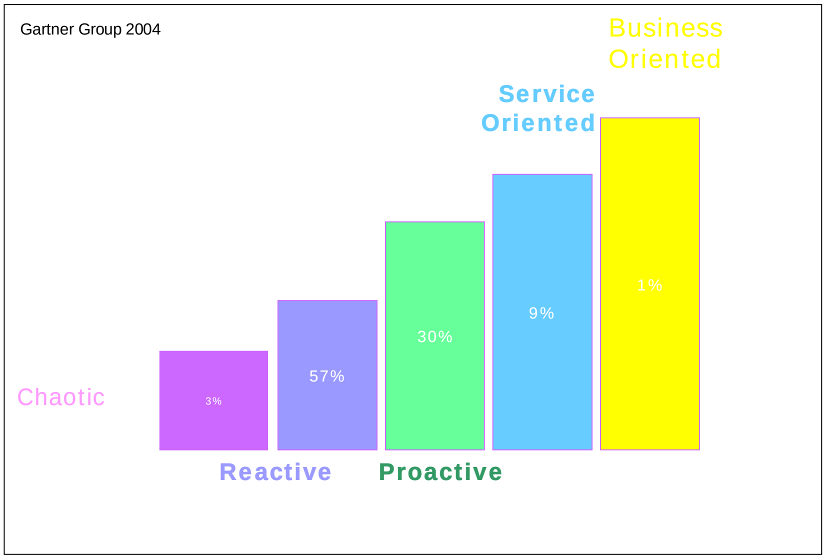


Figure 9: Survey (2004) of IT Capability Maturity Status

Capabilities Maturity Models were first proposed as a method of enhancing the quality of software products in the 1990's by the Software Engineering Institute at Carnegie Mellon University. These "best-practices" and well defined organization procedures are credited with significant improvements in software quality. Since then they have been adapted to other disciplines such as Systems Engineering and IT. Currently there is no single well accepted standard for IT although several are proposed, Gartner Group (consultancy group), Sun Microsystems (network equipment vendor) and CIBIT/Vrjie University (consultancy/academia alliance).

As network management has become more focused on best-practices and demanded better performance metrics from vendors they are beginning to be proactive in modifying the network as performance degradation is building and before it becomes service affecting. Moving forward, this expertise in the capabilities of the network is being viewed as having the potential for introducing new revenue generating projects to the organization while the operations of the network becomes more automated or self-adaptive.

Management of a self-adaptive network is driving research into how to distribute the intelligence from a centralized Network Operations Center (NOC) to a distributed intelligence with in the network itself. This in turn creates a shift in the role of the NOC from device management to providing network policies which support SLAs and enterprise initiatives without neglecting basic application traffic. This is true management of the network, rather than the devices and requires communication information management (CIM). Standards are being developed, CIM/DEN [39, 40] which focus on the network as a complex information system managed by policies that support information flow and not just data flow.

As this trend continues it will become increasingly important to monitor application specific data flow and performance rather than the flow of all aggregate traffic. This presents both a need and an opportunity. It is focus that has inspired my research project.

2.2. Simulation and Modeling to Support Network Research

Using modeling and simulation for network research addresses the issue of conducting experiments in a live network, which cannot be permitted due to potentially negative impacts on users. In addition, because a live network is not a controlled, except in a general way, or well characterized application environment, experiments to isolate the attribute under investigation run the risk of over-simplification or impacts from unanticipated behaviors.

Today’s networks are large and varied by most measures, e.g. geographical dispersion, number of nodes, link capacity (bandwidth), number of interacting protocols, types of applications, etc. This increase in complexity gives rise to unforeseen interactions (unpredictable by purely mathematical models of the individual components) as packets traverse the network from source to destination. Researchers have been measuring LAN, WAN, and Internet traffic and have discovered phenomena referred to as emergent behaviors. One example of this is the self-similar scaling, or fractal like, behavior of Internet traffic over multiple time scales. Emergent behaviors are a characteristic of complex systems. To understand these emergent phenomena two types of models exist [24]: evocative and explanatory.

Explanatory models have a mathematical basis which leads to the prediction of real world phenomena which can be sought for validation of the model. Evocative models demonstrate mechanisms by which observed phenomena can be produced based on empirical data rather than a deeper analytical theory. These are useful in analyzing the impact of change within limited bounds. Both are valuable in advancing network research.

This research captures the complexity of two semi-autonomous networks, i.e. managed by two IT departments connecting and coordinated by a campus-wide backbone and IT department. They have distinct topologies, one building verses two, and number and configuration of switches and routers. The traffic generated in the two networks is distinct both in volume and behavior over time. To this real-world topology and traffic pattern is added well-controlled, well-characterized application traffic. This gives an explanatory model representative of a single vendor university, or corporate intranet. It is appropriate for an investigation into the utility of the proposed indicator for application performance monitoring and management.

2.2.1. Modeling and Simulation Overview

Law and Kelton [41] are the established source for modeling and simulation foundations giving basic definitions of systems, models, simulations, and the basic variables of a simulation.

This section discusses only those most pertinent to telecommunications network modeling and simulation.

“A *system* is defined to be a collection of entities… that act and interact together toward the accomplishment of some logical end [… proposed by [42])”. The nature of the ‘logical end’ has a large influence on the level of detail and components of the system and how they are actually modeled as in the following example. A *model* can be either physical or analytical, that is, mathematical. If the “logical end” of the analysis is for facilities management to understand how many switches can fit in an equipment closet, cardboard physical models of the equipment might be appropriate. If, on the other hand, the amount of traffic a switch could handle is the “logical end” then the model in (1) would be more useful.

$$\begin{gathered} \lambda = average\_input\_traffic\_load \\ \tau = average\_output\_traffic \\ \textbf{average time in switch} \\ \mathrm{T} = \tau - \lambda \end{gathered} \tag{1}$$

The logical objective of telecommunications systems is to transport analog (audio, video, images) and digital (data) information between applications at the end nodes. Today virtually all transmissions are digital, usually converted from analog to digital at the end nodes. These bits are modeled according to the conventions of the various entities. Application entities transmit messages on an internally predictable but seemingly random basis. Protocols fragment the messages into conveniently sized packets, which include not only the application information but also transmission directions to get the packet to its destination, with quality and priority instructions. These units of transmission, their generation, and transformation are the network *traffic* the basic interaction between entities in the topology.

For communications network simulations the system ***entities*** are the information, digitized as bits, the transmission media (links) between workstations/telephone hand sets (end nodes), hubs, switches, routers, and other transmission and mediation devices, each with a variety of transmission protocols, buffers, and queues (resources). In an object oriented system these are modeled as attributes of the entity. These nodes and links represent the network *topology*. The basic functionality of theses entities is buffering, queuing and servicing the traffic in the system. Servicing the information bits is moving them from source to destination

The type of ***simulation*** can be classified along three axes:

- Static (e.g., Markov) or dynamic (evolving over time)
- Deterministic (system of equations) or stochastic (probabilistic inputs)
- Continuous (state constantly changes) or discrete (state changes at points in time)

Dynamic, Stochastic, Discrete event simulations with a warm-up period, are the most precise for packet level studies but the computational costs are high, consequently time to execute is high [43].

Flow models, sometimes referred to as fluid models, aggregate the events, tracking them as averages based on analytical models. These models are suitable for capturing steady state behavior and are typically used to evaluate traffic as source/destination bit streams[44].

Hybrid systems, where some aspects of the network are modeled using flow models and others are pure packet level event driven, now dominate. Both research simulation packages like NS2 and commercial packages like OPNET are typically hybrid tools[45].

When analyzing the output statistics from a steady state simulation consideration must be given to the validity of the statistics, i.e. they must not include values from the warm up period. In addition, if a long run is necessary to the objectives of the study, analysis of the results of the several runs needed to validate a stochastic simulation (calculation of the confidence intervals, variance, etc.) becomes tedious. Fortunately output can be taken from a long terminating or steady state simulation by collecting statistics in batch intervals and using the batch means rather than several runs to eliminate statistical bias.

Finally there is a distinction to be made about whether a system or a simulation is terminating or steady-state [46]. This distinction influences startup conditions and statistics collection. For a system that is idle, then running, then idle again there are clear startup and termination states in the system and it is obvious what is needed in the simulation. Telecommunication networks are ‘always on’ which might argue for steady-state simulation. However for many network investigations a terminating simulation can be used if the simulation goes through a warm-up period to insure that it has reached a steady state, i.e. routing tables, queues, buffers, etc. are in a representative state. For example traffic engineering is frequently concerned with peak hour conditions. If this is the objective then, after a warm up period, to reach peak conditions, statistics can be gathered and the simulation can be terminated as appropriate. If the objective is an analysis of the transitions from peak to slack than a long but still terminating simulation can be anticipated.

2.2.2. Strengths and Weaknesses of Simulation for Network Research

Because data communications networks have become so integral to society (for example in e-commerce, e-learning, e-government, etc.) operational, or live, networks can not serve as a test-bed for well controlled research experiments. The application of simulation technologies to network research is well understood[47], as well as the caveats associated with using this research technique. Simulation has been used extensively in network research, in areas a diverse as the performance of protocols [12, 48, 49] and architectures [50-52], analysis of routing [48, 53, 54]and other [1, 55] communications algorithms, and for self-adapting networks to predict performance for decision analysis by Intelligent Agents [32, 38, 56]. Simulation is also used extensively in industry for network planning and optimization[45, 57].

Recently the credibility of network research using simulation has been questioned due to discrepancies between research and real-world experience as well as an inability to reproduce the results of the research. Frequently this has been a failure to manage the complexities of simulation, for example, recording the decisions made in configuring the models or in setting up the simulation runs and/or collecting and analyzing results. Pawlikowski performed one of the first surveys of published network simulation research, 2200 articles in IEEE publications, and states the problem as follows:

> "… we have witnessed another success of modern science and technology: the emergence of wonderful and powerful tools for exploring and predicting behavior of such complex, stochastic, dynamic systems as telecommunications networks. … this enthusiasm is not shared by all researchers in this area. An opinion is spreading that one cannot rely on the majority of the published results on performance evaluation studies of telecommunications networks based on stochastic simulation, since they lack credibility." [58]

This concern was further analyzed by surveys of published research in mobile ad-hoc networking (MANET) [59, 60] where Kowalski et al analyzed papers from the ACM International Symposium, MobiHoc 2000-2005. Simulations supported the results in 114 of the 151 papers. These papers were evaluated against four areas that lend credibility to research results: it's repeatable by other researchers; the simulation is unbiased due to initialization bias or inappropriate pseudo-random number generator (PRNG); it's rigorous, in designing scenarios that actually test the question being investigated and multiple scenarios to avoid selecting a single, special case scenario; finally it's statistically sound in data collection and data analysis, for example giving confidence intervals relative to data points. The percentage of papers that met the criteria established in each of the four areas was in the teens or lower. For example, only 7% addressed initialization bias and none mentioned the PRNG used leaving open to question what bias might have been introduced by the simulation technique.

While both papers above provide good descriptions of the pitfalls to be avoided when using simulations in network research the most recent (July 2006) article on simulation credibility from the MANET research community [60] summarizes and amplifies the discussion and offers it’s similar but more complete list of recommendation. For this dissertation they are generalized for network management research and will be discussed in Chapter 5.

This highlights some of the challenges in simulating communications networks: Accurately modeling even small networks, appropriately tuning parameters, providing details of the simulation setup and a rigorous analysis of results.

2.2.3. OPNET (Optimized Network Engineering Tools) for Research

With commercially available DES tools it is possible to model a campus intranet in sufficient detail and of sufficient complexity to research application traffic performance in a meaningful and in a broadly applicable way for Network Management research.

Researchers have validated simulated network behavior against the behavior of real network traffic. One very pertinent study [61] compared FTP and CBR (constant bit rate) traffic in a laboratory test-bed (5 nodes) to NS2 and OPNET simulations of that network to validate the fidelity of the simulations to the live network. They found pros and cons with both simulations and difficulty with setting up a valid comparison between the three networks, one real, two simulated. Their conclusion after tuning all systems was that the tuned OPNET simulation was slightly more accurate and easier to work with overall although for the specific CBR test NS2 was slightly more accurate.

2.3. End-User and Quality of Experience (QOE)

Many disciplines are concerned with the End-User experience and have used various terms to describe this area. Frequently, it is overloaded on the term QOS [10] by the network management community and includes network deployment or provisioning as well as customer support. QOS is also favored by the computer science community [17] when designing distributed application although it is generally a subset of application performance issues. In this research the term QOS is limited to the metrics collected at the nodes to gauge network performance. QOE is favored by the infrastructure vendors [14] and the Human Computer Interaction community [13] and described below as it is used in this research.

2.3.1. Basic Concept

Application sensitivity in monitoring the network is becoming more significant as networks become a commodity critical to society rather than an engineering feat. One contribution of this research will be to provide an application sensitive detector of performance degradation to enable the NOC or the IA to detect and remedy performance problems directly at the level where it occurs before the customer registers a complaint.

This dissertation examines measuring network quality of service (QOS) from an application-centric rather than a data transmission perspective a notion referred to as quality of experience (QOE). As shown in Figure 10 some of the research is in the domain of human factors, some in computer science and distributed computing, and some in the IT organizational development. This research considers only the performance management of the network and the contribution it makes to the overall QOE.

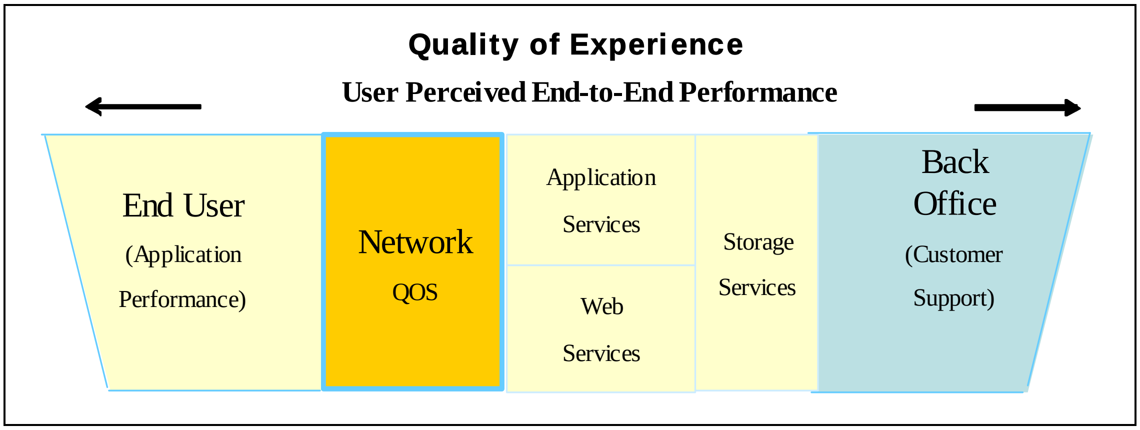


Figure 10: Full Span of Quality of Experience

This requires modeling performance at a higher level of abstraction than the current QOS values such as capacity, queue length, bit error rate, latency, etc. there is a whole matrix of values that fully specify an application. In research related to network design for future multichannel systems [62] expands the application specification for video on demand (VOD) to three distinct environments: Application Server, Network and End-User Device.

$$VoD_{spec} = QS_{QoS} + QS_{QoN} + QS_{QoD}$$

$$QS_{QoS} = \{< framerate, [5\,fps..40\,fps] >, < colordepth, [2bit..24bit] >, < resolution, [320x200;800x600;1024x768] >\}$$

$$\mathbf{QS_{QoN} = \{< bandwidth,\ [10Kbps..512Kbps] >, < loss,\ [0.0\%..0.01\%] >, < latency,\ [5ms..500ms] >\}}$$

$$QS_{QoD} = \{< colors, [2..24bit] >, < NetworkInterface, [802.11b;802.3] >\}$$

Figure 11: Fully Specified Application

As more application designers shift to fully specifying the network performance requirements network management can shift from managing device performance by QOS values to managing customer satisfaction with classes of applications using the current QOS values compared against the required values. Most research in this area has been focused on measuring performance of infrastructure components while industry researches the appropriate uses of these metrics in Service Level Agreements (SLA).

2.3.2. Current QOE Research

There are several approaches to this research area and the focus here is from the network management and the end-user perspective. This is not to minimize the strides being made by computer scientists and software engineers who are developing architectures and algorithms to minimize the impact of common network degradation on the end-users. Their efforts are supported by the research of both HCI and network research and their success contributes to making the experience of both network operations and end-users more satisfying.

Currently some network researchers are evaluating existing quality of service metrics, already well established in deployed networks, and how these measures relate to the business of providing multimedia services. It should be remembered that it was primarily voice and data transfer applications the drove the initial QOS metrics. Today, and into the future, network traffic and network quality will be more dynamic and complex. Businesses will be organized around either understanding the mechanisms that need to be implemented in network devices to guarantee QOS for all bits/bytes or packets of traffic or conversely understand the nature of the traffic and develop flexible adaptive systems [62-64] that can adapt to new application requirements and still fit into existing infrastructure.

For VOIP Beuran et al[63] proposed a QOE metric , although they have named it User Perceived Quality (UPQ), that uses PESQ an implementation of the original subjective voice quality measure Mean Opinion Score (MOS) but this will only serve for voice applications. For data transfers, round trip time and packet loss are considered to derive proposed Goodput and Transfer Time Performance (TTP). Siller and Woods [13], proposed that with a sufficient understanding of the traffic arbitration could be used to optimize the QOE using select QOS metrics. It would be advantageous to be able to do one calculation, perhaps based on Lp-norm, which could be weighted for each class of traffic.

This would be supported by the HCI research which can serve to tie the network conditions to the user experience. Due to the distributed nature of massively multiplayer online games (MMOG) network and HCI researchers have joined together to analyze traffic from first person shooter games[65], Madden NFL football[66], and Quake[67] while evaluating the experience of the end-user. With streaming video and video conferencing gaining popularity research has been conducted to evaluate the application performance under known network conditions and the subjective impact on the end-user[68, 69]. With more research available developing the weights for the proposed QOE indicators is feasible.

2.3.3. Challenges facing QOE as a Key Performance Indicator (KPI)

Today typically link capacity utilization, traffic demand as a percentage of link capacity, is the KPI monitored. Based on sufficient available capacity for all application traffic it is assumed that the network performance is adequate for each type of application.

To detect performance degradation the network operations center (NOC) primarily monitors average capacity utilization in the undifferentiated traffic. A highly researched area is distributed network management, using Intelligent Agents (IAs) throughout the network. Consequently, in anticipation of distributed network management, it is reasonable to monitor the network health at multiple levels. At the operations management level as a qualitative indicator of customer satisfaction this could trigger a deeper quantitative analysis resulting in corrective action.

CHAPTER 3: APPLICATION-CENTRIC KEY PEFORMANCE INDICATORS (KPI)

“Not everything that counts can be counted and
not everything that can be counted counts.”
A sign in the Princeton office of A. Einstein

3.1. Problem Statement

Until now it has been possible, and necessary as a foundation, for performance management to focus on each device or platform in the telecommunications network, the basic assumption being, if each device performed within desired specifications then the communications requirements of the application traffic would be satisfied. This leads to performance management rules-of-thumb such as “If link capacity utilization is less than 20%, e.g. 20Mbps in a 100Mbps link, congestion is unlikely and thus delay, packet loss and throughput requirements will be satisfactory”. This heuristic assumes that switch configuration has sufficient queuing and buffering for the traffic presented, that bandwidth shaping for the node is not unduly penalizing certain classes of critical applications, etc. Until customers complain, application performance is unknown at the Network Operations Center. Increasingly customer care is a revenue generating concern while the volume and variety of application placing different demands on the increasingly hybridized (by protocol such as TCP/IP over ATM, etc) or converged (by technology, wireline to/from wireless, etc.) are increasing the complexity of what must be managed. It may no longer be adequate to manage the network at the device level.

Network performance management, whether by an internal Information Technology (IT) department or by an external provider, is central to end user satisfaction with the communications network. Network performance measures commonly describe behavior, such as delay, loss, corruption, transfer speed/volume, etc., of undifferentiated streams of bits, byte or packet traffic measured at the nodes traversed in the network. Typically these measures are referred to as Quality of Service (QOS) metrics or parameters.

Distributed or client/server applications are designed to perform optimally within some specified range of values for key QOS parameters, i.e., those that have the greatest impact on that application's performance. To determine the importance and optimal range for QOS parametric values application designers must relate human computer interface research to the network performance measurements. Performance from the perspective of the application user is sometimes referred to as quality of experience (QOE). Once application designers specify the performance requirements which will insure customer satisfaction some department or business deploys the application under either an implicit or explicit service level agreement (SLA) with network management. These agreements, or performance expectations, are based on average network QOS values for all traffic from all applications using the network. Network operators, human or automated, typically monitor one or two performance figures of merit, frequently the upstream and downstream node throughput. Network customers receive reports (typically monthly or quarterly) to inform them of compliance with performance levels. Here the average values for all QOS parameters of interest, as well as security and reliability metrics, are reported in some detail. SLAs may include payment for a superior grade of network performance or penalties for failure to perform at some agreed upon level in these cases performance becomes a revenue concern.

For network management the challenges are to architect, design and deploy a network that meets the requirements of many customers each of whom will be running a variety of applications with more or less stringent network performance requirements and demands on the network resources. Performance needs are generally met by “over-provisioning”, i.e. deploying a network with greater capacity then required by the customers, at increased network costs which at some point may become uneconomical. Network planning is based on SLAs, implicit or explicit, and growth projections.

Once the network is deployed network operations must maintain performance for each customer based on device , such as switches and routers, QOS values which are an aggregate of all the application traffic for all customers through that node. This can lead to customer dissatisfaction since the averages may be within limits even though one application is consistently denied the desired performance. To add to the performance management challenge, customers may add entirely new applications to their traffic mix or upgrade to a new version which place greater demands on the network and/or an application may become wildly popular, for example the Napster music sharing application, making the growth projection obsolete.

As distributed applications, such as multimedia of all types, massively multiplayer online games, grid computing, and simulations, become more central to the fabric of global society it may be advantageous, even appropriate, for network performance management to measure and monitor some new application-centric quality of experience (QOE) parameters. This dissertation investigates the feasibility of evaluating the network QOS parameters against the requirements of application classes (data base queries, browser searches, voice and video over IP). This type of QOE metric has the most utility in a TCP/IP access network since it is one where the network provider and the application end-users have the most direct interaction. In addition access networks are typically TCP/IP where the basic philosophy is to network heterogeneous devices with best effort service, although now various means have been developed (bandwidth shaping, DiffServ, IntServ) to prioritize traffic. These QOE metrics require disaggregating network traffic into data flows of a single application type. This is being done today in “back room” near real time passive monitoring devices such as the Cisco Network Analysis Module (NAM).

Further, since an application’s performance is usually specified by several QOS parameters (such as throughput, delay, packet loss tolerances, etc.) a method for deriving a single measure of QOE for each application class would aid management when troubleshooting which applications are not receiving sufficient resources to keep their end-users satisfied. Here a mapping of the multiple QOS values to a good/poor performance QOE by application qualitative index becomes significant.

Finally, to simplify the network monitoring task, a high level indication of whether the node (switch, router, etc.) is performing satisfactorily or whether it needs to be investigated could be indicated by deriving a QOE metric for the node which maps the QOE application qualitative value (good / poor) for each application class present in the traffic stream to the relative importance of that application class, i.e. is it mission critical (VOIP) or background (email). This ranking and rating could be implicit, based on general business usage, or specified in a SLA for special cases where simple FTP is part of a critical interactive application. Thus the node QOE index would be customer specific.

The problem addressed in this chapter is the development of algorithms for indicators of end-user Quality of Experience (QOE), and building a simulation to demonstrate their utility for network performance management. Two indicators are proposed, QOE_{app} and QOE_{node} , and the rationale for the proposed algorithm discussed. The indicators are intended as a trigger for network operations management to optimization device configuration or create a new forecast for network expansion. The ability of these indicators to determine application and node performance, in qualitative yet actionable, terms can be demonstrated using simulation since more empirical data must be analyzed before a rigorous mathematical approach can be taken. In addition simulation is needed to develop the proof of concept for this approach before it would be implemented by equipment vendors. The simulation used to validate the QOE indicators will be described with its strengths and weaknesses. Finally, issues of data analysis, external validity, and limitations are examined.

3.2. Mathematical Rationale for QOE Indicators

While seeking a mathematical approach to this problem it is important to ground the mathematics in network realities. Based on insights from the network researchers cited in Chapter 2 the indicator should clearly be based on QOS values which are used as performance indicators today using a statistical process control approach. In addition applications are already designed to perform within certain QOS parameter limits, i.e. specified requirements (Table 2: Application QOS Specification) so these are two network realities to be included in a new application-centric approach.

Another driver for the qualitative approach is that, while the infrastructure may generate all QOS values, these are just raw data and not actionable. Networks operations are managed according to some selected QOS values, typically just the undifferentiated Throughput for each node (switch or router) or an indicator of capacity utilization for either the link or the node. If there is a mathematical approach that would integrate several raw data points this would make performance monitoring more effective.

The basic concepts behind the proposed application-centric performance indicators are:

a) there exists some function for each application class, be they current or future applications, that based on network performance as a function of the traditional QOS parameters describes optimal or good performance expected by the end-user. This function will be called QOE_{app}.

b) the performance of each node can be described by a function which integrates the performance of the application classes present at any time *t* , and maps that performance to the service level agreement, implicit or explicit, for the satisfactory performance of that segment of the network. This function will be called QOE_{node}

A promising area for the mathematics to underlie a new application-centric approach to network performance management is Functional Analysis (FA), a branch of mathematics concerned with the study of spaces of functions and coming from the calculus of variation it implies a function whose argument is a function. This dissertation will not attempt to explore the potential of FA, Banach Spaces, or L^p-norm in a rigorous mathematical manner, in fact in network research a more empirical approach may be needed due to emergent network phenomena. Consequently algorithms based on FA concepts, emphasizing network relationships, are proposed and modeled in a campus simulation to evaluate their utility.

Normed vector spaces are at the heart of FA. In this dissertation |X| will be used to denote the vector norm which is a quantity "that in some (possibly abstract) sense describes the length, size or extent of the vector". It is common in traffic management [70]to relate applications to a two dimensional measurement space (Figure 12) this dissertation proposes to extend this to a vector space of as many QOS metrics as an application need. The vector space must also be non-dimensional, i.e. not tied to and specific time or space measure, e.g. throughput in bits per second, or delay in milliseconds. This is achieved by populating the vectors with the abstraction %QOS (actual value / specified value).

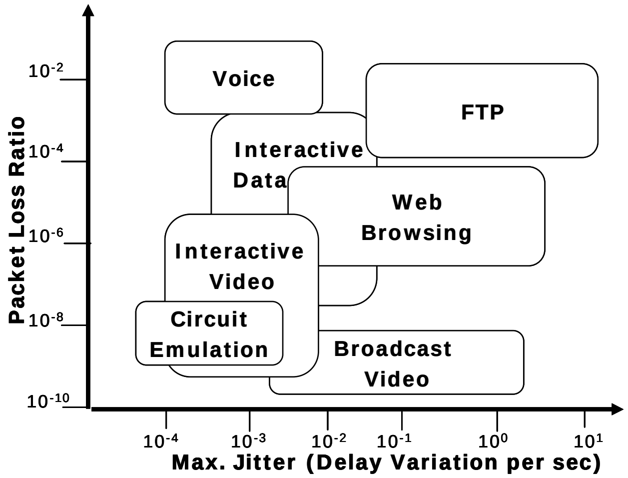


Figure 12: Application QOS Requirements for Parameters Packet Loss and Jitter

Thus at any time (*t*) the network performance can be quantified as a vector of QOS performance metrics $|QOS_{actual}|$ whose members represent parameters such as delay, jitter, lost packets, etc. Similarly the quantitative performance requirements of an application can be specified as QOS parameters and described as vector $|QOS_{spec}|$. This leads to the first, very simple function which gives a vector of quantitative performance measures which will be called %QOS.

$$\%QOSVoD = \left|(ActualDelay / SpecDelay), \quad (ActualPacketLoss / SpecPacketLoss) \quad , \quad (ActualThroughput / SpecThroughput)\right| \qquad (2)$$

In (2) the QOS parameters important for video on demand (see Figure 11) are calculated. Other parameters may be important for other applications for example both delay and jitter are important to Voice over IP, while FTP is only concerned when packet loss at a high level so the elements of %QOS vectors would be different giving:

$$\begin{aligned}\%QOS_{VoD} &= |\%QOS_{delay} \quad \%QOS_{loss} \quad \%QOS_{th}| \\ \%QOS_{VOIP} &= |\%QOS_{delay} \quad \%QOS_{jitter}| \\ \%QOS_{FTP} &= |\%QOS_{th}|\end{aligned} \tag{3}$$

The next step is transform these vectors to a single result, i.e. the QOE_{app} and to map the quantitative %QOS metrics to a qualitative index of good to poor performance from the end user perspective. The l^p-norm, is a generalization of the absolute value and is computed as in Equation 3. It is a power function in which each member of the vector follows a characteristic power curve. The final summation of all $|x_i|$ will also fall on some characteristics power curve.

$$|X|_p = \left(\sum_{i-n} |x_{i-n}|^p\right)^{1/p} \tag{4}$$

Using l^p-norm transforms offers several attractive characteristics; first, the summation allows for scaling and applying constants to the individual %QOS values; second, since the power (p) can be selected to map the values to a power curve characteristic, in a qualitative subjective way, to end-user perceived QOE_{app} performance (good to poor).

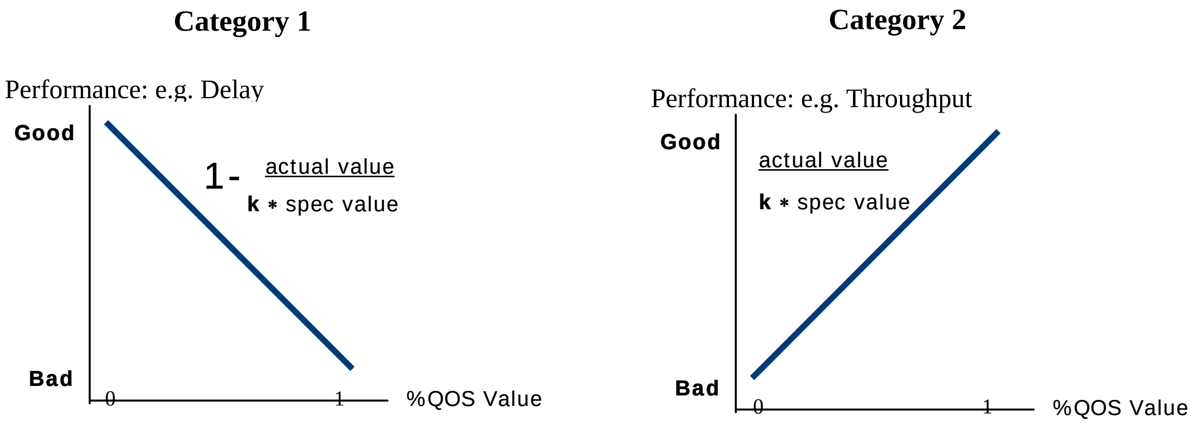


Figure 13: QOS metrics may have slope of -1 or +1 and scaling constants

For a satisfactory network performance level application developers and network management currently select an appropriate ATM, IntServ, or DiffServ service level based on the application categories in Appendix A. These are becoming too simplistic given the explosion of networked applications. One application may have tasks that are real time and other tasks that are interactive data.

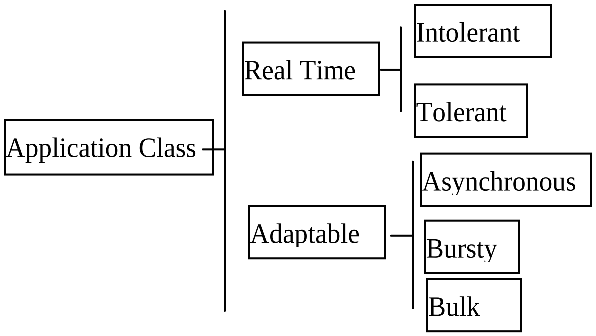


Figure 14: Taxonomy of Application QOS Classes

Rather than have the application developers select a pre-established network-centric performance service level, the next generation network providers should be able to take application specific requirements and monitor the performance of the application or the several application classes such as Data Base, HTTP, Video on Demand which are aspects of the same application. Since the QOS metrics are well understood and used by application developers as well as network management they should form the basis of the QOE indicator that would be monitored when network management shifts from network-centric to application-centric. Even before a thorough human factors analysis is done for each application class, an anecdotally based first approximation can be used to select a power for each application when calculating QOE_{app}.

For example, a power of 4 would indicate an application whose QOE_{app} is “good” (say >.85) when the raw %QOS is < 80%. Although it needs verification, this could be true of FTP since it is tolerant of most performance degradation. So to transform the %QOS vector for FTP traffic a p=4 would be used. On the other hand a very demanding application like VOIP would have a p=.25, etc. The research to map application %QOS to a characteristic power curve is outside the scope of this current effort. Based on simulated traffic power values will be selected for the applications used in this investigation. Partial list is given in Figure 15

$$|QOE_{app}\ |^{p} = \left(\sum_{i-n} |\%QOS_{i-n}|^{p} \right)^{1/p}$$

$$|QOE_{VoD}\ |^{p} = \left(\sum_{i-n} |\%QOS_{i-n}|^{2} \right)^{1/2}$$

$$|QOE_{VOIP}\ |^{p} = \left(\sum_{i-n} |\%QOS_{i-n}|^{..25} \right)^{1/..25} \ |$$

$$|QOE_{FTP}\ |^{p} = \left(\sum_{i-n} |\%QOS_{i-n}|^{4} \right)^{1/4}$$

Figure 15 Power Selection for Application Classes

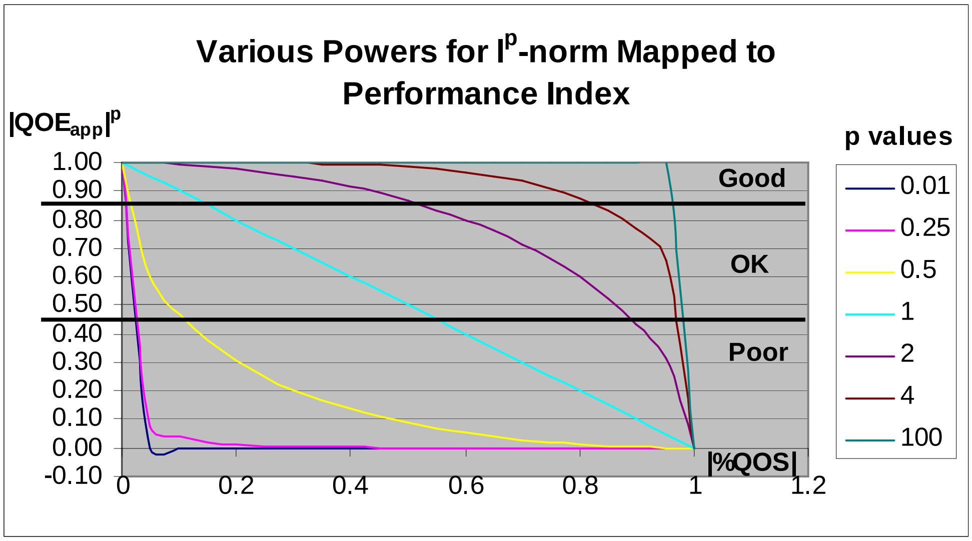
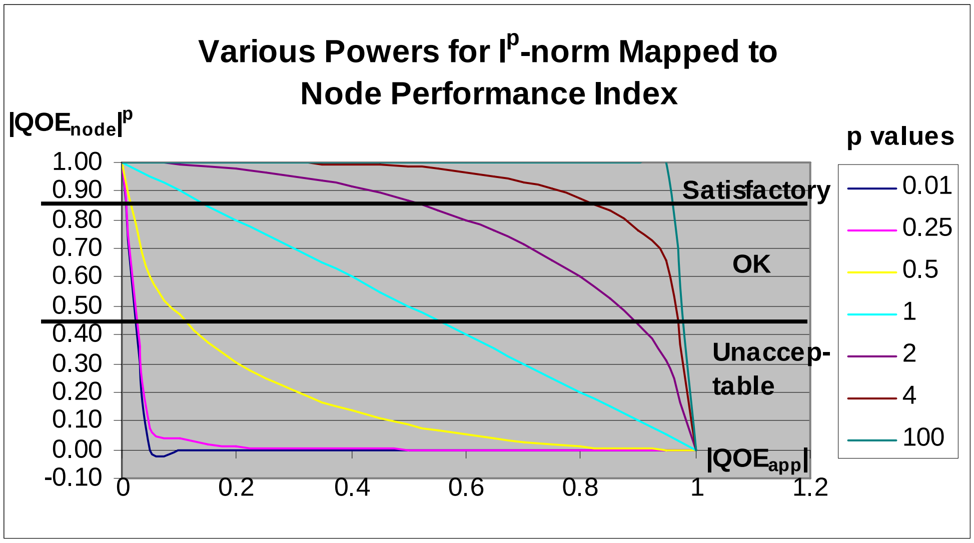


Figure 16 Power Curves and Performance Index

Further if the l^p-norm is applied to a vector | $\text{QOE}_{\text{app } i\text{-}n}$|, representing performance of the mix of application classes present at a node during any measurement period (t) a new qualitative value QOE_{node} results. Like QOE_{app} the qualitative index is tied to the power selected as characteristic of the desired node performance. The index would be organizationally specific. For this research the 0 to 1 values represent application traffic performance unsatisfactory to satisfactory for the mix of traffic at that node. This is an actionable indicator that could be used by network managers to analyze the need for troubleshooting, reconfiguring or optimizing the node.



The criteria for selecting a characteristic power curve would again be organizationally specific and represent the SLA the network management has with it’s customers. This research has used the criteria that the core nodes are of higher performance value due to the large number of end-users who would be affected by unsatisfactory performance and those nodes are measured against a more stringent curve, such as p=.25. On the other hand, local switches can tolerate greater end-user dissatisfaction because a smaller population is affected so they would be measured against the p=4 curve.

$$|X|_p = \left(\sum_i w |x_i|^p \right)^{1/p} \tag{5}$$

Another promising aspect of the l^p-norm approach is that it includes weighting factors, see (5, which allow the formula to weight QOE values for network and application considerations. For example $\mathrm{QOS}_{\mathrm{spec}}$ parameters may be weighted to indicate the relative importance of that parameter to the overall performance of the application. Again $\mathrm{QOE}_{\mathrm{app}}$ values may need to be weighted to indicate the %of over all traffic each application represents when calculating. $\mathrm{QOE}_{\mathrm{node}}$.

3.3. Simulation Test-bed to Demonstrate Effectiveness of QOE Indicator

A valid and credible simulation must be utilized to gain insight into the proper formulation of the QOE KPI since well controlled experiments cannot be performed in a live network. This dissertation utilizes simulation to demonstrate the ability of the proposed Quality of Experience (QOE) key performance indicator (KPI) to inform network management of the quality of customer experience on a per node basis given some mix of application traffic. Experiments to vary both the volume of traffic and the proportion of the different application classes within that traffic are needed will give a preliminary indication of the utility of these KPI.

The results of these simulation experiments (scenarios) will allow a comparison between the expected application performance built into the scenario, that is, a known satisfactory experience vs. an intentionally unsatisfactory experience. The values reported by the QOE indicators will be compared to the existing QOS metrics to determine if the application-centric KPI are reliable indications of application performance at each network node (switch or router).

In the real world this could result in network management proactively addressing and optimizing performance before the customer complains.

Modern telecommunications simulation environments, such as NS2 or OPNET, are open source allowing the researchers to modify standard device models to evaluate modifications to the protocols or device performance. There are standard statistics collected in this protocol/device models and custom statistics can be added. The QOE indicators discussed in this chapter will be implemented as a custom statistic and displayed and stored during the simulation runs.

3.3.1. OPNET Simulation Environment

The modeling and simulation tool selected is OPNET (OPtimized Network Engineering Tool) an open source, object oriented, commercial product with a dominant position in both government and industry network planning and research. The strengths and limitations of both the simulation tool and the network models are discussed in Appendix B.

The perspective of this dissertation is both transport layer traffic and the network management overview. Undertaking a model and simulation of a campus size network requires a full featured and robust simulation tool rather than programming extensions to a highly efficient but limited package. The criterion for a simulation tool reflects the two perspectives and their demands. For protocol development there is little need to have application specific traffic sources a stream of bits is sufficient. Typically they are abstracted. Due to the application-centric nature of this research those were important criteria. For the scenarios proposed in the case study the ability to capture and import actual samples of futuristic applications was also significant. Open source is needed so the proposed KPI can be implemented and statistics gathered on its performance. The criteria used are summarized below:

- Ability to model a variety of protocols and transport media (wireline and wireless)
- Models of Cisco equipment (layer 2 and 3 switches primarily)
- Hybrid tool to allow some traffic to be modeled as flows and others packet level
- Tool stability and support
- Wide recognition and acceptance of the tool
- Flexibility to extend models and custom statistics
- User-friendly input and output to deal with the magnitude of the problem space

The speed of execution is not one of the criteria and early indications are that simulation speed will be the main trade off required to accomplish this research. Appendix B summarizes the tool evaluation.

3.3.2. Network Topology

The topology of this simulation is based on the network and traffic of a research institute, Institute for Simulation and Training, and one of the science colleges, the College of Optics and Photonics, connected to the campus backbone of a metropolitan, multi-campus, state research university, the University of Central Florida.

Networks are frequently viewed as either “core” or backbone, where large volumes of packets are transported at high speeds for long distances for example WANs or the Internet, or “edge” or access points, where users connect to the network and applications generate traffic, for example LANs or client / server database applications. There appears to be a cycle in network research

Viewing the Internet as a network of autonomous networks, there is a mix of every hardware and software available, that is, it is a heterogeneous network in every way. This makes it difficult to model the dynamics of such a large complex system (>100k nodes). Small networks (>5 <100 nodes) are typically used for network research leading to unanticipated scalability issues. This research explores modeling a midsize network (~50k nodes) to use as a test-bed for network manageability research. Viewing the modeled network as a socio-technical network focuses attention on the application specific nature of the traffic.

The complexity of these segments is deemed sufficient to permit interesting experiments inducing traffic, equipment failures, and protocol interactions to degrade application performance which should be detected by the QOE KPI.

3.4. QOE Hypothesis and Simulation Assumptions

As mentioned in the problem statement this research hypothesizes that functions exist that will describe network performance in terms meaningful to application user quality of experience. Although there is not yet enough empirical data to derive the differential equations or functional analysis equations to describe this function mathematically, a beginning can be made using simulation science.

Algorithms have been developed, inspired by l^p-norm and firmly grounded in network realities, to serve as key performance indicators (KPI) of application-centric network performance. These are qualitative and actionable values calculated on a per node, rather than end-to-end, basis. These algorithms will be developed in Excel® and may be implemented in the simulation time permitting

To demonstrate the effectiveness of these KPI a campus access network will be simulated. For the purpose of establishing a power assignment for the QOE_{app} calculation the premium applications will be VOIP and Video Conferencing. Web and data base traffic will be generated at a mid-priority. Low priority or "best effort" applications will be email and FTP. For the QOE_{node} a simple value of service will be established based on the centrality of the node, that is, a node in the backbone will have a high service priority, core nodes or nodes serving a building will have medium level of service priority, while poor QOE at the very edge of the network will be more tolerable since fewer customers will be affected. Statistics will be gathered from the simulation and exported to spreadsheet for further calculation of QOE or other non-standard metrics.

The access network as model for this research simulation represents the complexity of real networks in these ways:

1. The topology and configuration of the switches and routers was imported from the real architecture (e.g. star topology, VLAN configuration) and design decisions used for the campus network. It is not an unrealistically simple topology
2. Campus traffic was collected for a week, using the program MRTG, and used to create a background traffic load on the appropriate link. This accurately models the aggregate, uncharacterized by application, traffic produced by the end-users of these switches. It varies dramatically by day of the week and hour of the day along the same link and even more from switch to switch i.e. link to link. To use just a few, or worse yet a single, probability distribution function (PDF) would be a gross oversimplification.

3. To have some disaggregated application traffic to allow QOEapp calculation, standard models of application traffic available in OPNET were used. These traffic generators use appropriate PDFs for message size and transmission time. In addition, application appropriate packet size and, in the case of a VOIP, appropriate codec processing.
4. Traffic volume is varied in the simulation by using the LAN model and varying the number of users and the applications they use.

While this simulation does not represent the Internet complexity it does represent the access portion of the Internet where the service providers, either ISPs or Intranet providers interact most directly with the network user.

Link utilization statistics, the current network management tool, will be monitored as for comparison with the proposed QOE KPI.

3.5. QOE_{app} and QOE_{node} Algorithms for Simulation

These are the steps to be executed in either Excel or within the OPNET simulation as a custom statistic:

QOE_{app}

1- Declare application QOS parameter specification constants for each application, e.g.

VOIP - throughput = , jitter= ,

FTP - throughput = , dropped packets=

2- Declare power constant for each application based on application user's tolerance to network fluctuation. The same value will be used for each QOS parameter as well as the application as a whole

3- Collect the statistics of interest from the simulation and modify as needed, e.g. collect node throughput and subtract from link throughput rating to derive available throughput

4- Calculate %QOS for each application present during that period,

e.g. ((link throughput capacity - current throughput) / app throughput requirement) $^{\text{app power}}$ = %QOS$_{\text{thru}}$

(jitter / (app jitter requirement / # hops)) $^{\text{app power}}$.=. %QOS$_{\text{jitter}}$

5- Calculate QOE$_{\text{app}}$ for each application present during that period,

e.g. (%QOS$_{\text{thru}}$ + %QOS$_{\text{jitter}}$ % + etc) $^{\text{1/ app power}}$ = QOE$_{\text{app1}}$

(%QOS$_{\text{thru}}$ + %QOS$_{\text{jitter}}$ % + etc) $^{\text{1/ app power}}$ = QOE$_{\text{app2}}$

The value obtained is then evaluated against the QOE$_{\text{app}}$ index, an abstraction based on human factors research, to establish whether the application performance is good or poor. There will be challenges in determining if the simulation produces the necessary statistics on a per node basis to describe the application performance. In addition a meaningful means of establishing how application requirements for end-to-end performance can be treated for node statistics. The initial approach will be to divide the end-to-end specification by the number of hops.

QOE$_{\text{node}}$

1- Calculate QOE$_{\text{node}}$ for the mix of applications during that period,

e.g. (QOE$_{\text{app1}}^{\text{node power}}$ + QOE$_{\text{app2}}^{\text{node power}}$ + etc) $^{\text{1/ node power}}$ = QOE$_{\text{node\_ABC}}$

The value obtained is then evaluated against the QOE$_{\text{node}}$ index, determined by the network provider, to establish whether the performance is satisfactory or unsatisfactory.

3.6. QOE Test Simulation Scenarios

Several scenarios have been designed to test the ability of these QOE KPI to detect poor network performance based on an application-centric view of the data plane. Many scenarios will need to be evaluated varying all the independent variables. In this research the independent variables are considered to be:

- Network Provisioning: designed and deployed capacity of the nodes (switches and routers) and links to process and transport traffic could be varied. This would study the value of the QOE indicators when capacity is constrained due to economics or as traffic increases.
- Application Mix: which applications with their characteristic interactions, message length, and performance requirements could be varied. This would study the value of QOE indicators when new applications are introduced or when one application has an unanticipated increase in usage changing the demands on the network resources.
- Protocol Mix: a variety of protocols such as TCP/IP with or without QOS provisions, ATM, TCP/IP over ATM or wireless or optical switches could be model to study the QOE indicators in different contexts.
- Device Mix: a heterogeneous network with equipment from a variety of vendors would introduce variations in processing capacity, scheduling and queuing algorithms, protocol stacks etc. This would study the sensitivity of the QOE indicators to these minor differences in performance.

For this first investigation only network provisioning will be varied all other independent variable will be held constant. The dependent variables will be the QOS metrics, throughput, delay, jitter, and dropped packets and the QOE indicators calculated from them.

3.6.1. Scenario 1 – Over-provisioned

This scenario will have the baseline topology and link load imported from the campus as the aggregate traffic modeling the real world mix of applications, simple TCP/IP protocol as implemented in Cisco equipment. To this is added an overlay of disaggregated application traffic from which application specific QOS metrics can be collected and used in the QOE calculations. The expectation is that all QOE indicators will show all applications performing well and each node giving satisfactory performance.

3.6.2. Scenario 2 – Under-provisioned Links

This scenario will have all the characteristics of scenario 1 except that the links between nodes will be reduced in capacity by an order of magnitude. The main QOS metrics that are expected to be affected are throughput and delay, perhaps jitter. The expectation is that QOE indicators will show some applications performing well but other, more sensitive applications, performing poorly. Similarly nodes at the edge might give satisfactory performance based on their lower traffic load and service priority while the core and backbone nodes might reach an unsatisfactory level of performance.

3.6.3. Scenario 3 – Under-provisioned Nodes

This scenario will have all the characteristics of scenario 1 except that the internal traffic processing capacity of the node will be reduced by an order of magnitude. The main QOS metrics that are expected to be affected are dropped packets, delay, perhaps jitter. The expectation is that again QOE indicators will show some applications performing well but other, more sensitive applications, performing poorly. Again nodes at the edge might give satisfactory performance while the core and backbone nodes might reach an unsatisfactory level of performance..

CHAPTER 4: EXPERIMENTAL RESULTS USING QOE KPI

> *"To know what you know and what you do not know: that is true knowledge.."*
>
> K'ung-fu-tze (Confucius)

The results fall naturally into three categories and are reported in those terms.

- Design of a credible and suitable simulation test-bed (Verification and Validation)
- Formulation of QOE key performance indicators (KPI)
- Evaluation of QOE key performance indicators (KPI) in various scenarios

Due to recent challenges to the credibility of simulations used in network research and in the hope that other researchers may wish to use the test-bed for other experiments, its development is reported in detail. Similarly with the formulation of the QOE KPI, they are currently first approximation to an approach for transitioning network management from network-centric to application-centric.

4.1. Simulation Test-bed

To achieve a representative level of complexity and realism the campus network of a metropolitan research university was used as a proxy for any campus or corporate intranet with similar characteristics. The topology (nodes and links) represents three buildings with primarily research users and was established using Cisco DCI configuration files. The traffic from the live system was collected in early June 2006 using the Multi Router Traffic Grapher (MRTG) tool. The OPNET multi-vendor import (MVI) module, made available to this research under a special limited use license, made the importation of real-world topology and traffic possible. The campus network operations center (NOC) collected the data and reviewed its use under a network security agreement for researchers (see Appendix C). The use of the MVI module is straight forward and well documented but as always there were issues getting the correct data from the network in a form that could be imported. Without MVI a network simulation of this size and detail would have been very difficult to create[71].

Recently, due to an inability to reproduce the results of other researchers as well as discrepancies between research and real-world experience which may be inevitable given the current understanding of network performance, the credibility of network research using simulation has been questioned. Since a simulation is the least expensive, and in cases where the protocol, equipment, or metric under investigation doesn’t exist yet, it’s the only way, to perform early network research this concern must be addressed. The credibility concern may be due partly to emergent behavior in the real network or simply a failure to adequately manage the complexities of simulation i.e. configuration control and statistical validity of stochastic results.

Pawlikowski [72] performed one of the first (2002) surveys of published network simulation results, analyzing 2200 articles in IEEE publications. This work was continued with two surveys of published mobile ad-hoc networking (MANET) [59] [60] research. These papers highlight some of the challenges in simulating telecommunications networks:

1. accurately modeling even small networks
2. having an suitable pseudo random number generator (PRNG)
3. appropriately tuning device parameters/ configuration
4. providing details of the simulation setup
5. a rigorous analysis of results

The results of designing and developing a credible simulation test-bed are reported here focus on its impact on the QOE metrics research.

4.1.1. Topology

The campus, in common with many corporate intranets, has made the decision to select a single vendor, Cisco, for economy of scale. Again like many other networks they have solved the problem of performance with over capacity (over provisioning). All links typically run at 5% or less of capacity since a common rule-of-thumb for planning holds that congestion will degrade performance if traffic is routinely over 20% capacity, especially with TCP/IP networks.

For investigating the utility of the QOE indicator two autonomous networks, each with its own IT department, were modeled. One network is in two buildings giving a total of three networks, A, B, and C. The core of the campus network was simplified to model just the paths connecting the three networks. Two Cisco layer-3 switches form a backbone for all the main campus and connect to remote campuses. Remote campuses, although modeled during the MVI import, were all remove. Although network-A connects directly to the backbone, networks-B and C connect through other switches designated Core-*switch_model*. This research does not investigate the performance of the QOE at the network core so this decision seems justified. The simulation test-bed consists of 29 layer-2 switches, 3 layer-3 switch/routers with 40 gigabit links connecting them. The work stations and departmental LANs that form the network below the switch level are not modeled explicitly rather the network traffic they generate was captured and modeled as link load. Based on the switch interface information it is estimated that there are greater than 1100 nodes below the switch level in networks A, B and C. This information is summarized in Figure 17: Network Topology

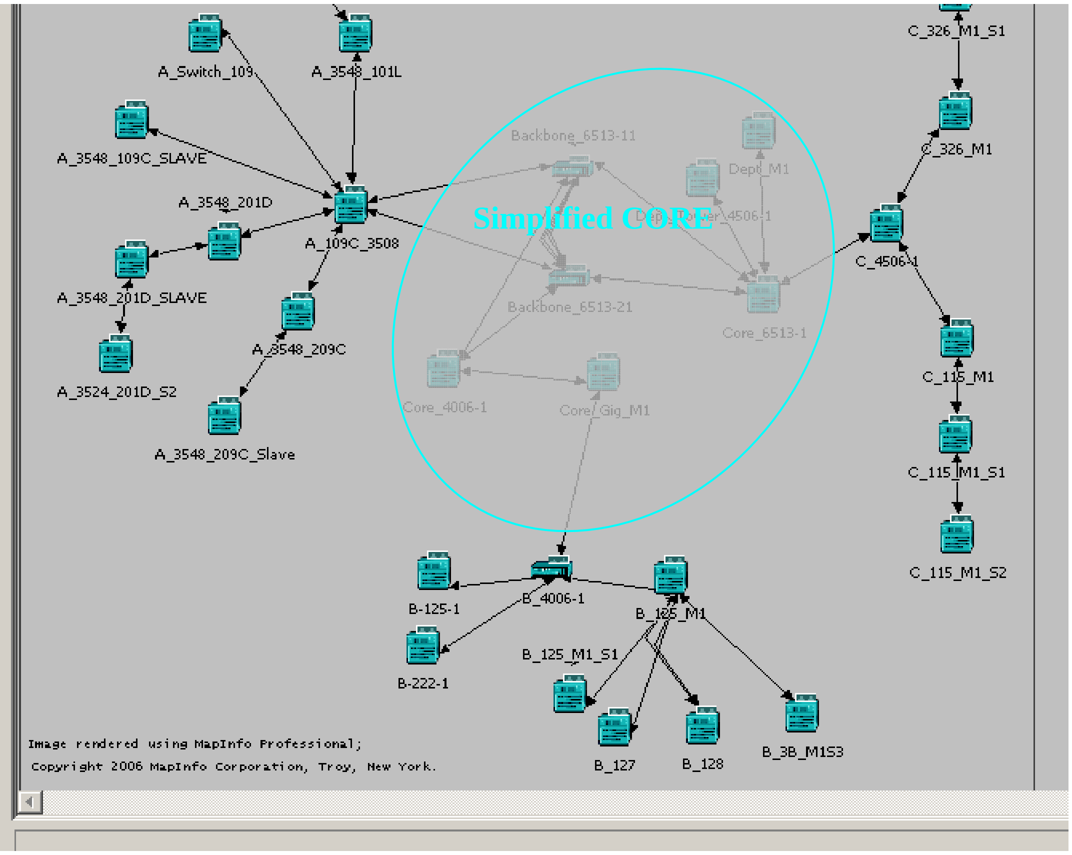
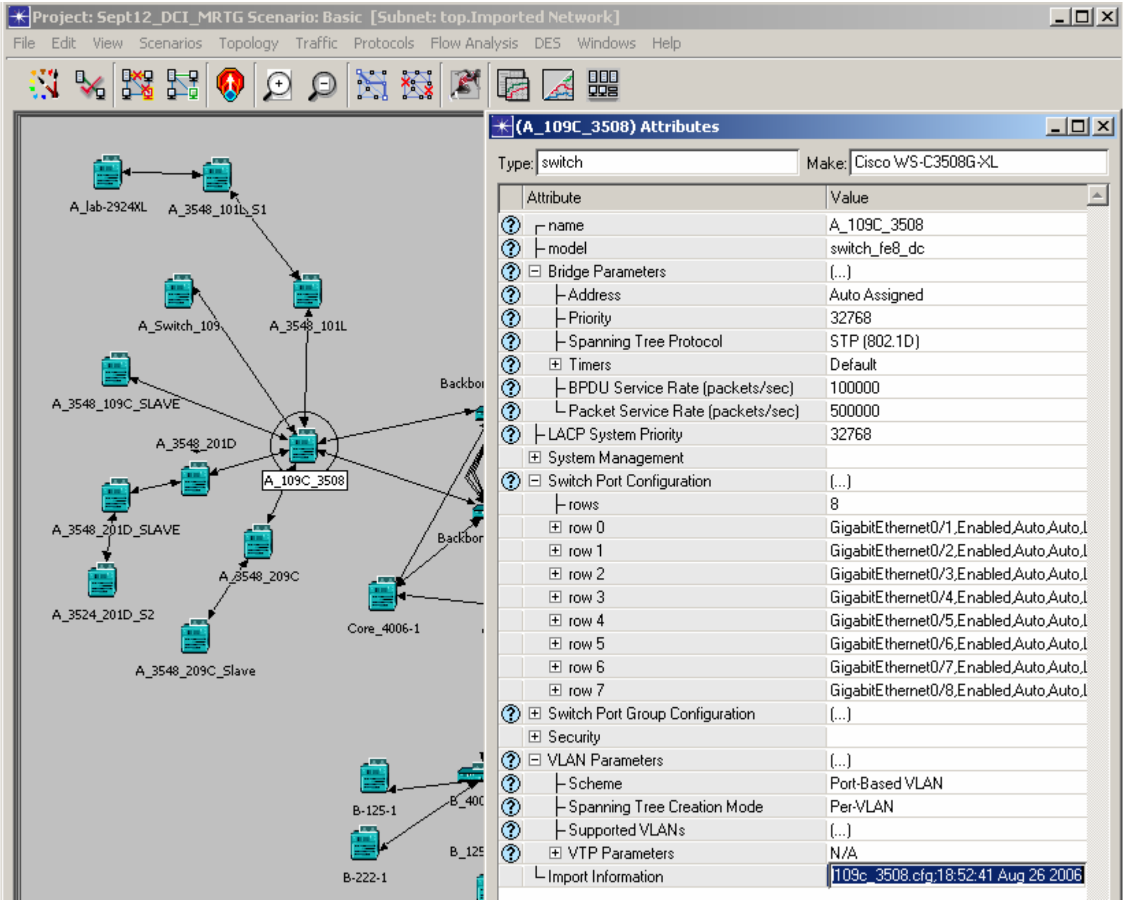


Figure 17: Network Topology

While OPNET is open source and allows the lower layers of the model to be modified when researching a new protocol or routing algorithm, the attribute editor GUI gave sufficient control for this network management investigation. Figure 18 Complex Switch Parameters Set from Imported Configuration) illustrates the nested abstraction of attributes for one of the switches.



Type: switch Make: Cisco WS-C3508G-XL

Attribute	Value
name	A_109C_3508
model	switch_fe8_dc
⊟ Bridge Parameters	(...)
Address	Auto Assigned
Priority	32768
Spanning Tree Protocol	STP (802.1D)
⊞ Timers	Default
BPDU Service Rate (packets/sec)	100000
Packet Service Rate (packets/sec)	500000
LACP System Priority	32768
⊞ System Management	
⊟ Switch Port Configuration	(...)
rows	8
⊞ row 0	GigabitEthernet0/1,Enabled,Auto,Auto,
⊞ row 1	GigabitEthernet0/2,Enabled,Auto,Auto,
⊞ row 2	GigabitEthernet0/3,Enabled,Auto,Auto,
⊞ row 3	GigabitEthernet0/4,Enabled,Auto,Auto,
⊞ row 4	GigabitEthernet0/5,Enabled,Auto,Auto,
⊞ row 5	GigabitEthernet0/6,Enabled,Auto,Auto,
⊞ row 6	GigabitEthernet0/7,Enabled,Auto,Auto,
⊞ row 7	GigabitEthernet0/8,Enabled,Auto,Auto,
⊞ Switch Port Group Configuration	(...)
⊞ Security	
⊟ VLAN Parameters	(...)
Scheme	Port-Based VLAN
Spanning Tree Creation Mode	Per-VLAN
Supported VLANs	(...)
⊞ VTP Parameters	N/A
Import Information	109c_3508.cfg;18:52:41 Aug 26 2006

Figure 18 Complex Switch Parameters Set from Imported Configuration

To indicate that an attribute is an abstraction and has sub-attributes a + is placed in front of it. To indicate a list of values (…) is used. Here the attribute ‘Switch Port Configuration’ is expanded as well as the ‘VLAN Parameters’. There are 8 ports configured on this switch and some of the sub-attributes are shown on the right. The VLAN attribute ‘Supported VLANs’ is a simple list of VLANs and is indicated by (…). Given the complexity of the models the ability to import these settings from the live system is, at the very least, useful! The basic features of the topology were validated through discussions with the network operations center (NOC).

4.1.2. *Traffic*

The use of empirical traffic streams rather then strictly generating traffic from a distribution allowed the simulation greater variability in link utilization with opportunities for real world unforeseen emergent phenomena. While traffic per se was not under investigation it seemed a good approach for this test-bed simulation. MTRG data for seven days in June 2006 were imported using OPNET-MVI. The data shows typical bursty-ness and diurnal fluctuation.

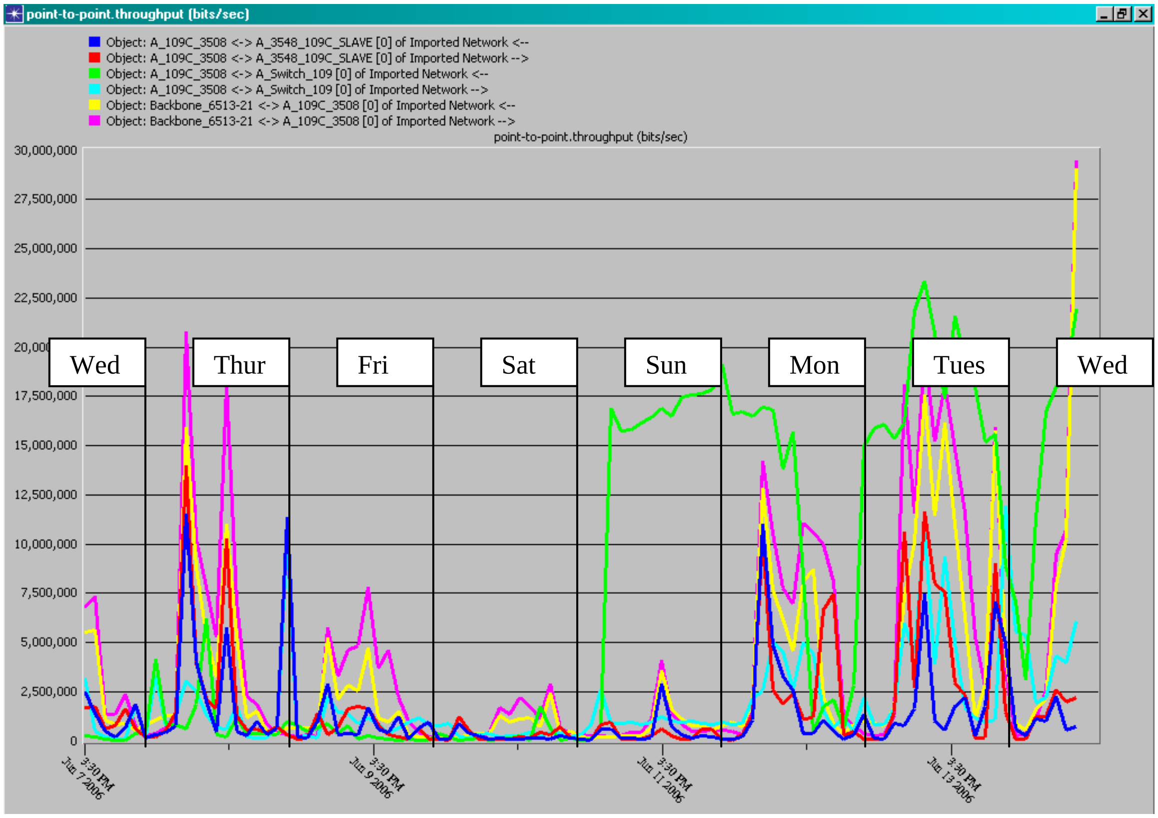
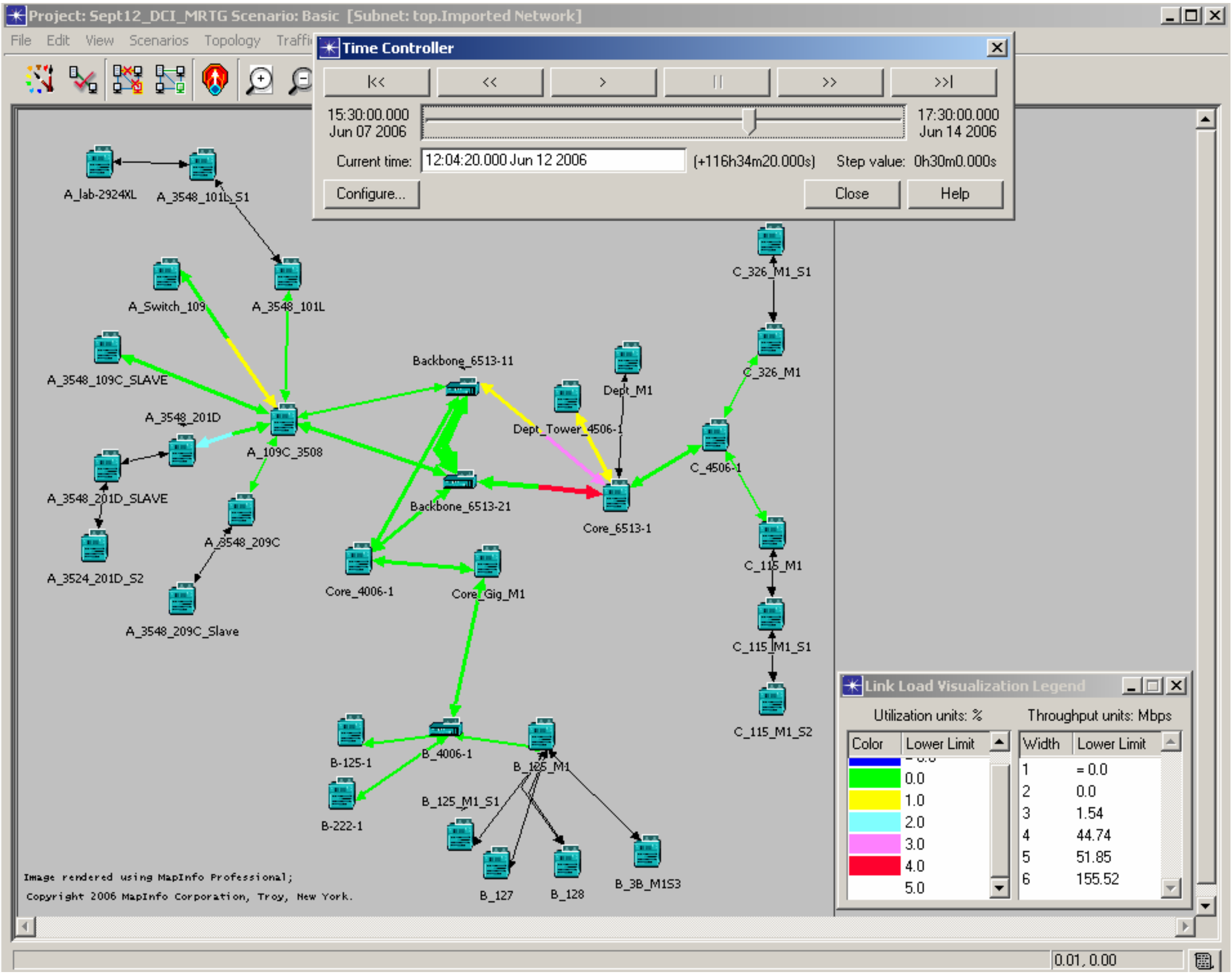


Figure 19 (Uplink / Downlink Traffic for 3 Links)

After the network traffic was imported inspection of the model showed that data for the lower level of switches, i.e. those furthest from the backbone, had not been captured.



15:30:00.000 Jun 07 2006 — 17:30:00.000 Jun 14 2006

Current time: 12:04:20.000 Jun 12 2006 (+116h34m20.000s) Step value: 0h30m0.000s

Utilization units: %

Color	Lower Limit
🟦	= 0.0
🟩	0.0
🟨	1.0
▨	2.0
🟪	3.0
🟥	4.0
⬜	5.0

Throughput units: Mbps

Width	Lower Limit
1	= 0.0
2	0.0
3	1.54
4	44.74
5	51.85
6	155.52

Figure 20 Traffic Visualization

Using the Time Controller and Link Load Visualization features, as shown in Figure 20 Traffic Visualization, make this obvious. Rather than ask for more MRTG data to be collected, an empirical traffic distribution from another link was selected by traffic volume and based on being unrelated to the current traffic path to avoid biasing the statistics.

The basic real-work topology and traffic models used in the development of test scenarios is shown in Figure 21 (Basic+Lower Link Traffic). The links are color coded links by peak utilization with most links utilized to only 1% or 2% of capacity. This campus is an intentionally over-provisioned network to insure sufficient bandwidth for a growing population and increasingly complex applications.

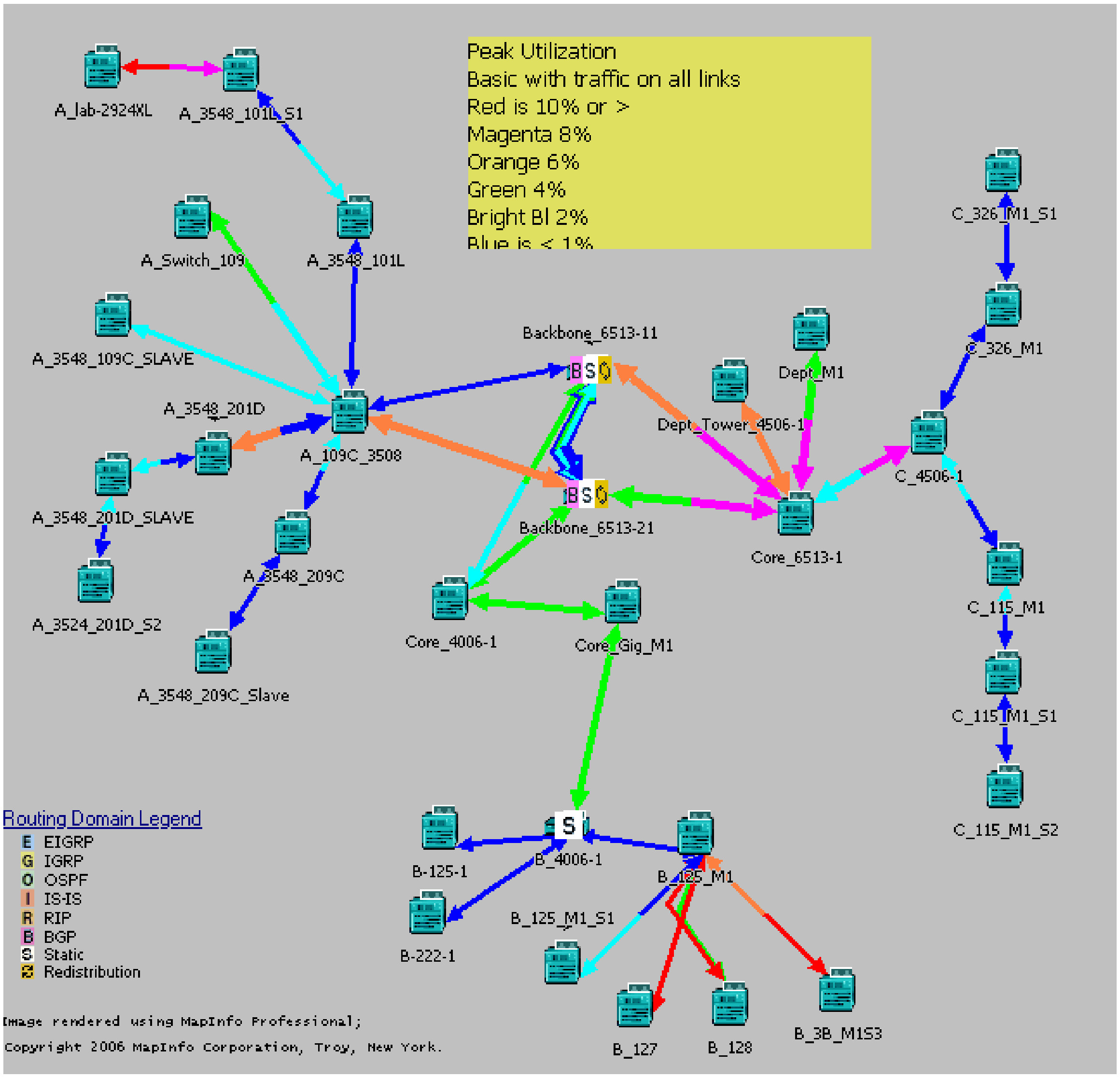


Figure 21 (Basic+Lower Link Traffic)

4.1.3. Simulation Runtime Setup

After modeling a representative network the next consideration is configuring the parameters of the simulation itself so that the results are valid and appropriate for the experiments being run.

As a general rule the OPNET default values were used for preliminary simulation runs to investigate the QOE KPI formulation. For the final evaluation some of these setting were modified and changes will be noted in the scenario.

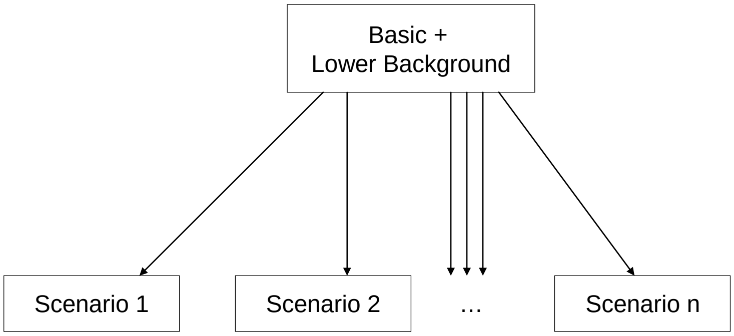
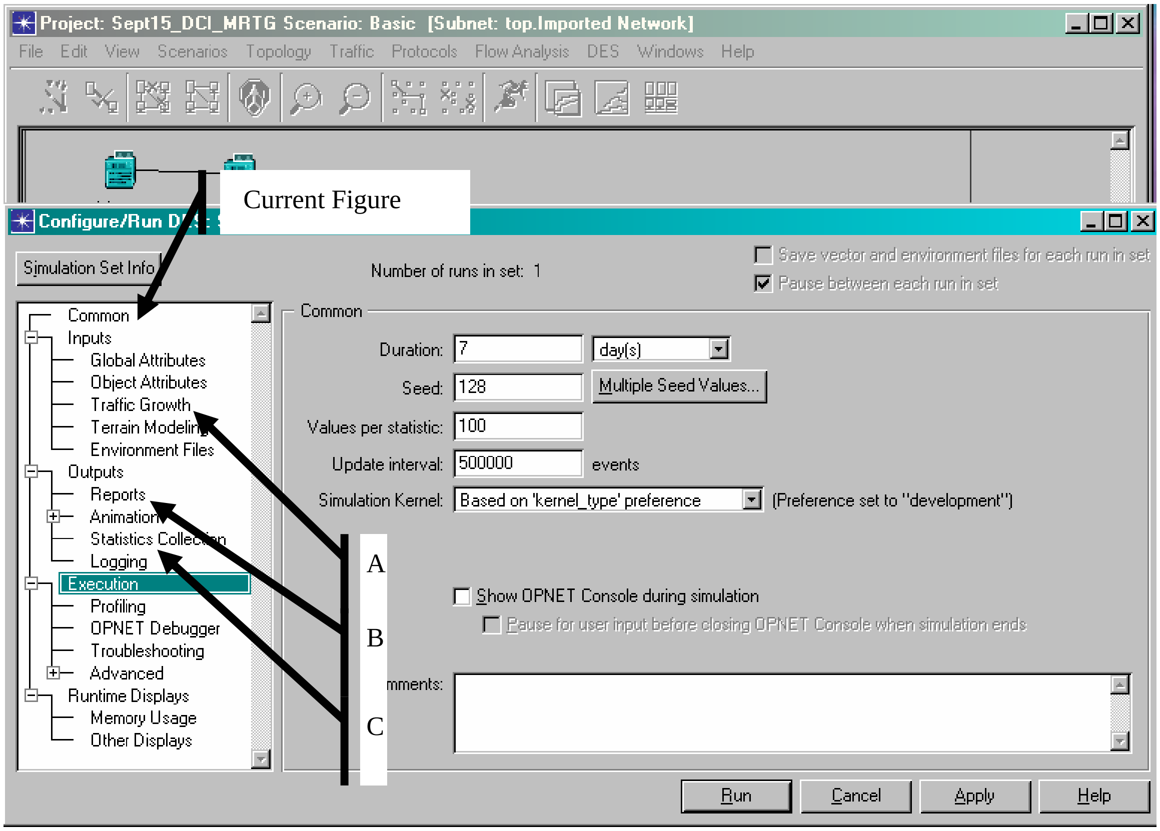


Figure 22 (Configuration Control)

First and foremost good software engineering configuration control must be practice throughout the project. The conventions established for this research are to separate the Basic-Test-bed as described above from any experimental scenarios. This resulted in the file structure in Figure 22 (Configuration Control). Within the project file of each experimental scenario OPNET offers a ‘Duplicate Scenario’ function to allow changes to the independent variables of the simulation and comparison of results. In this research the “over capacity” scenario was duplicated and used to create the “under capacity” scenario. The results of multiple scenarios can be statistically compared within the OPNET project graphically or in tables. As was done for the

QOE evaluation the results can be exported to comma delimited files for import to Excel or other programs and manipulated further.

Once the basic simulation project management is set up each run of the simulation has to be configured so the results are statistically valid within the basic limitations of simulation research.


Project: Sept15_DCI_MRTG Scenario: Basic [Subnet: top.Imported Network]

Configure/Run DES

Number of runs in set: 1

☐ Save vector and environment files for each run in set
☑ Pause between each run in set

Common

Duration: 7 day(s)
Seed: 128
Values per statistic: 100
Update interval: 500000 events
Simulation Kernel: Based on 'kernel_type' preference (Preference set to "development")

☐ Show OPNET Console during simulation
☐ Pause for user input before closing OPNET Console when simulation ends

Figure 23 (Configuration of Simulation Run)

The researcher has to make choices about the duration of the run to achieve the goals of the research. Duration of the simulation is selected based on the purpose of the study and in addition the statistics errors that can be introduced when the traffic has a high standard deviation due to the normal diurnal effects. Although there are 7 days of background traffic in the test-bed these experiments were run on peak hour replications which is also negatively impacted by the day to day fluctuation of traffic. This was not an important aspect of the QOE utility problem so one peak hour was chosen after experimenting with various duration settings. The settings available under A in Figure 23 (Configuration of Simulation Run) allow for selection of different time periods to be selected and if desired analytically increased. This was not an option used.

The next item on the common configuration panel is “Seed” for the pseudorandom number generator (PRNG) which is always a key consideration in stochastic simulations. OPNET the OS function call to random() which is typically a BSD (Berkley Software Distribution) version of the linear congruential generator algorithm good for over 2 billion numbers before the series repeats. Using this type of PNRG autocorrelation problems are avoided since in a long run (over 2 billion random events) will find the network in a very different state. For statistical validity of the stochastic results one run is never enough and several runs with different seeds given to the PNRG product statistically valid results. Five replications where run using the seed values: 128, 65, 193, 258, and 311.

“Values per statistic” and another panel “C: Statistics Collection” must be set to avoid initialization bias and too few values collected for statistical validity. Since this dissertation uses a terminating not steady state simulation bias is removed by not starting application traffic until the routing tables and link loads have stabilized. This is controlled by attributes of the applications and warnings occur during the run and collected in the DES log if background traffic hasn’t stabilized. On “C” the statistics collection panel statistics collection can be delayed. In these simulations I used a 5 minute delay before starting to collect statistics.

Finally item “B: Reports” must be set to specify what statistics will be collected during the run. By default no statistics are collected and the simulation is very fast but useless.

Specifying a statistic when it is unavailable, ATM statistics in a TCP/IP network does no harm but does increase the run time.

4.1.4. Credibility of Test-bed

OPNET® v11.5 was chosen for this research since it is a commercial product and an industry leader for telecommunication simulation. New releases occur at approximately 6 month intervals with enhancements and bug fixes. During the 10 month development of the test-bed it was upgraded from v10.5 to v11.0 and finally v11.5. Two newer versions have been released but the test-bed was not up graded.

The library of vendor device and application models, built up by OPNET, accelerates test-bed creation. An open-source philosophy provides object attributes and process editors for modification of device configuration and process, represented as finite state model, so performance impacting abstraction or refinement of device or application can occur. The simulation engine or kernel is proprietary.

Cost of commercial products frequently drives researchers to freeware like NS2. Fortunately, OPNET® has a generous university research program and my fellowship was able to fund a year of customer support. OPNET® Modeler is a hybrid DES package that includes various modeling techniques, primarily analytical (mathematical) and discrete event. OPNET® refers to a hybrid of the two as "micro-simulation"[73]. OPNET® also has a product, Multi-Vendor Import, (MVI) for importing switch and router configuration files as well as importing MRTG traffic data which can be used to load the links. Although this is not a normal part of the university program, a very temporary license was granted for this research.

The ability of OPNET® MVI to import real device configurations and network traffic adds accuracy to the model. The difficulty here is real-world campus network security. The campus Information Resource Management (IRM) group required a Network Security Researcher’s Agreement before collection, cleaning and allowing use of campus configuration or traffic data. Over time the information will be less security sensitive but until then researchers have to work under IRM guidelines which will control the configuration and traffic data used in research.

By utilizing a leading commercial simulation there is some assurance that the internals of the finite state machines, the pseudo random number generator (PRNG), etc. are being tested in real-world telecommunications development environments and are constantly being challenged and improved. For this research, wherever possible, standard OPNET® models and statistical measures were used. This will allow other investigators to understand the distributions and configuration parameters used by examining the v11.5 version of the application, statistic, or device process. Except for the LANs all other IP addresses were set to let the simulation auto-assign them. This led to uncovering a bug as will be discussed below.

4.1.5. Verification and Validation

In this dissertation validation refers to the level of correspondence between the model and the real world. Verification refers to a simulation that behaves as expected. Finally, it must be verified that the test-bed will run the desired experimental scenarios that will test the QOE metrics.

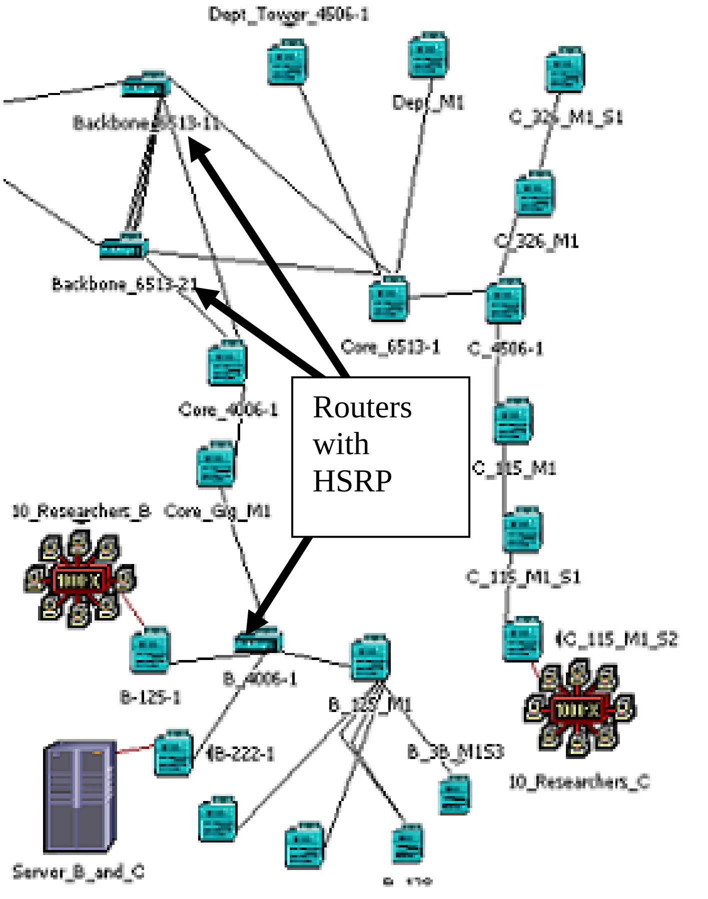
4.1.5.1. Validation

Since the validation was more straight-forward it will be discussed first. One of the advantages of importing the campus network was that the Network Operations Center (NOC) provided files and expert evaluation. Since no import process is fool proof, they compared my simulation topology (links and nodes) and traffic statistics to their averages and helped me debug some errors in the import. Since only a portion of the campus network was used there were many missing port connections. The import software sorted it out but with over 100 import error messages. After several attempts a model was arrived at which the NOC engineers agreed was a fair simulation of the three buildings and that the remaining error messages related to the backbone and core switches and missing portions of the campus. These errors increase simulation execution time but should not impact the validity of the statistical results.

4.1.5.2. Verification

Verification for this test-bed was far more complex and extensive. Since experimental scenarios were to be built by duplicating and modifying the baseline, a series of test simulations had to be run and statistics analyzed to verify that the application traffic from the LANs and the traffic routing exhibited the expected behavior in the baseline. Just one example of an issue that led to some very strange statistics was a misconfiguration of the application server for networks B & C, see Figure 24. The simplest test was from a Network B LAN, modeling the users, to the Server located in the same network. Traffic was sent by the LAN and as expected the same volume was received by the Server. Surprisingly, the Server traffic sent was larger than the traffic received at the LAN. Some other traffic was being sent by the Server elsewhere in the network. By sending my project to OPNET® Support they realized that I had configured the server attributes so it was also a client, thus generating traffic. A case of “you don’t know what you don’t know”.

Once traffic was sane within network B the next verification test was that the volume of traffic at the Server should equal the sum of traffic from both B and C networks. When the statistics between B and C showed otherwise it took the help of the OPNET® support center to determine that there was actually a bug in the OPNET® model of the Cisco hot standby routing protocol (HSRP) see Figure 24



Routers with HSRP

Figure 24 Networks B and C blocked by bug in OPNET

Problems occur when most but not all IP addresses are auto-assigned. A work-around using the VLANs was suggested until the bug could be fixed in an upcoming release. While it is inevitable that a complex model has bugs, the ability to have a support group, with access to developers, evaluate the model saves the researcher from becoming mired in coding and debugging a complex software environment.

While this major issue was being resolved the behavior of the various applications email, FTP, HTTP browsing, HTTP images, video conferencing, and VOIP from several LANs were evaluate for conformance to intersession and packet size probability distribution functions (PDF). The standard models seemed adequate for the purposes of this research and to facilitate the efforts of other researchers to reproduce the results.

4.1.5.3. QOE experiments

To develop the proposed QOE indicators many, many scenarios have to be run to analyze the utility of the proposed metrics under a variety of conditions. Consequently time characterizing and setting up the test-bed is time well spent. Experimental scenarios would be created by varying the independent variables and refining the formulation of the metrics based on analysis. In addition, since the simulation is stochastic, i.e. probabilistic, in nature and a goodly number of probability distribution functions are used in application traffic generation, care must be taken to use a statistically valid set of QOS values for the dependent variables on which the QOE formulation is based. Further care must be taken that the PRNG produces suitably long stream of random numbers so that the simulation can run can to completion without reuse of the random number series.

For this research the results of five runs, with different seeds to produce different random number streams from the PRNG, were averaged to arrive at a statistically valid set of QOS values to use in the calculation of QOEapp. Within each run a value was collected every ten seconds of simulated time giving 90 values for a 15minute run. The QOE indicator could either be calculated on an average of the 90 values to represent the whole 15minute period or calculated for each of the 90 ten second intervals to evaluate the sensitivity over time. The 15 minute interval is a trade-off between statistical validity, the time it takes to run the simulation, and what is representative of the real world.

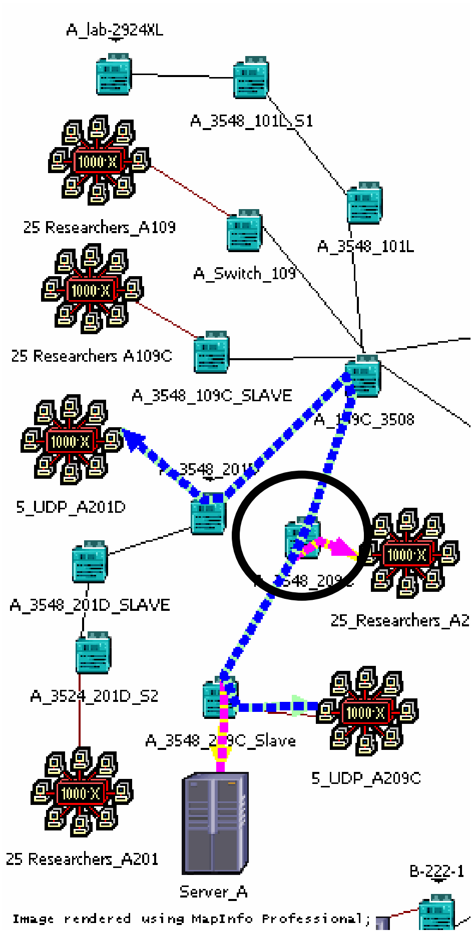


Figure 25 Focus for Statistics in 3 Scenarios

In the real world node statistics are collected at intervals that don’t add an undue traffic burden to the network yet allow evaluation of current network performance. Switch system management statistics collection can be set to 15 minute intervals, 30 minutes or 1 hour are more common.

In the simulation, as the number of events increases, for example as retransmissions occur as in scenarios where links or nodes are under-provisioned, the time it takes to run the simulation increases. Figure 25 represents a model which takes about 30 minutes to simulate a 15 minute period when running the baseline over-provisioned, scenario, i.e. the no problem model. For the more complex under-provisioned scenarios the same configuration takes an hour and a half since many more events are generated.

Finally data analysis and incorporation of proposed QOE statistics into the simulation must be addressed in a consistent fashion across all the simulations and runs. OPNET® provides an Analysis Configuration editor where vector panels and scalar panels can be created and turned into templates. Later they can be loaded with current data. Vector values represent the discrete values created over time. Figure 26 and Figure 27 are vector panels comparing values of interest across the 3 scenarios. A scalar panel will average the data from 5 runs into the single value shown here.

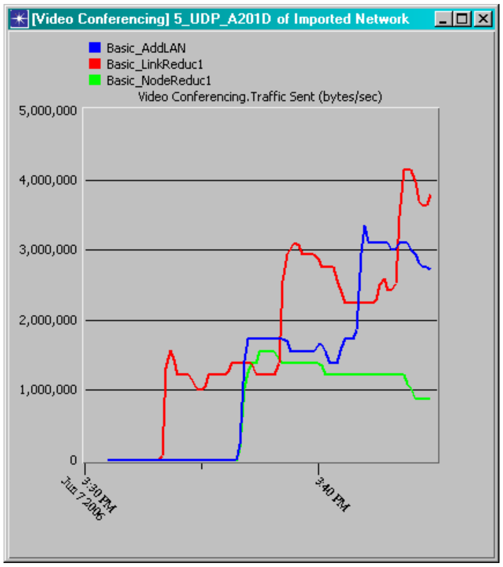


Figure 26 Video Conferencing Traffic in 3 Scenarios

Since the proposed QOE statistics are application-centric indicators of application user satisfaction, results from the test-bed have to be used to determined whether or not current network conditions, as simulated, will keep the customer satisfied.

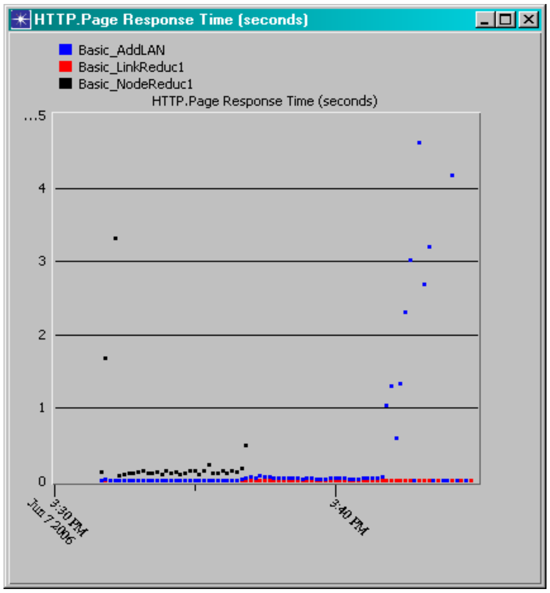


Figure 27 HTTP Delay in 3 Scenarios

This evaluation has to be by application class to determine if any one class is affected more than others. As Figure 27 shows it does affect applications such as HTTP, FTP, etc. so the question for this researcher is: “given the device-centric QOS values reported by the simulation for various application classes, do the new QOE metrics give network management a better indication of how degraded network performance is affecting the customer?” Let the investigation begin! The QOE performance indicators will be discussed in detail in future publications.

Finally, the actual incorporation of the QOE calculations into the simulation must be addressed. Custom statistics can be coded in C or C++, compiled and incorporated into the simulation so that all switches and routers could report their QOE results. Figure 28 shows the processes modeled for the switches and routers. The red arrows represent statistics interrupts from the kernel to collect values according to the scheduled collection period. In the campus network there are no system management statistics collected, hence the unconnected sysmgt icon above the switch icon. It would be reasonable to modify the test-bed to schedule a statistics interrupt and execute QOE statistics gathering at that point.

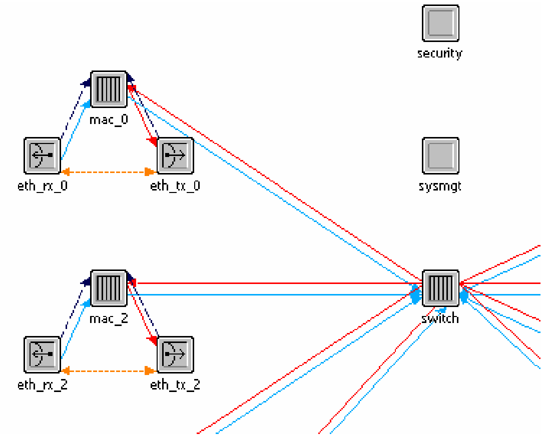


Figure 28 Switch Process Model

This was not done, since the QOE calculations are still in a very preliminary form and thus not yet worth the programming effort. However there are many paths to reach a goal in OPNET®. All statistics needed to calculate QOE where collected in one vector panel, which was turned into a template. Then the "export to spreadsheet" option was able to put them in Excel®. Within the spreadsheet many little adjustments could be made such as changing the throughput statistic into available throughput. The down side of this brute force method is that it limits the number of switches that can be evaluated since it is labor intensive. In Figure 25 the switch chosen for QOE evaluation is circled. This switch is of interest since all TCP based application traffic passes through it on the way to the Server (indicated by a magenta path). The two UDP applications, Video Conferencing and VOIP, LANs reach each other through this same switch (indicated by the blue path).

For the rest of the research I used OPNET Modeler v11.5.A PL3 under the University Research program. The OS was WinXP on two different machines. Changing machines only impacted the speed of execution since RAM was 1Gbyte in both machines, the Dell Duel Core 3GHz x 2 Pentium desktop configuration did not perform as well as the HP Athlon 3200+ in a laptop.

"A steep learning curve" is one of the complaints about OPNET and, while such a rich environment can never avoid complexity, the quality of the documentation, the customer support and the university liaison person make it a powerful research environment. Topology and traffic characteristics will be described below.

4.2. Formulation of QOE KPI

This section discusses the formulation of QOEapp and QOE node as related to network and simulation test-bed statistics. First approximations for formulae follow:

To populate the %QOS vector

$$\forall \quad \text{QOS requirements} \quad \%\text{QOS} = \left({\text{Current Value} - \text{SpecifiedValue}} \middle/ {\text{Specified Value}} \right)$$

$$\text{OR} \qquad (6)$$

$$\%\text{QOS} = 1 - \left({\text{Current Value} - \text{SpecifiedValue}} \middle/ {\text{Specified Value}} \right)$$

This is the link between the application development community and network management and the challenge here is to determine which of the statistics collected at the node (i.e. Current Value of QOS metric) best corresponds to the application's end-2-end requirement. This will be discussed in the results of the scenarios below.

Equation (7) is the first approximation of an application Quality of Experience metric. It is inspired by functional analysis and the lp-norm space which should permit a more definitive formulation as more empirical studies clarify the relationships between the nodal metrics and the end-2-end metrics.

$$\text{QOEapp} = \left(\left(\sum_{i-n} \left| \%\,\text{QOS}_{i-n} \right|^{p} \right) \middle/ n \right)^{1/p} \qquad (7)$$

Determination of the p value is based on the sensitivity of the application design, for example, if smoothing algorithms have been implemented in an application, such as VOIP or Streaming Audio, to reduce the impact of delay on audio quality as experienced by the end-user then the p assigned to that QOEapp may indicate less sensitivity to network performance degradation. For these initial experiments the p value is taken from the index used in [74] to determine bandwidth shaping. Thus p=.2 is the most sensitive and assigned to VOIP for these experiments. At the other end of the sensitivity scale is FTP with a p=5 indicating a very high tolerance for degraded QOS metrics.

While equation (8) is listed for completeness it was not part of this initial investigation for reasons discussed in the scenario results.

$$\mathrm{QOEnode} = \left(\left(\sum_{i-n} \left| \mathrm{QOEapp}_{i-n} \right|^{q} \right) \Big/ n \right)^{1/q} \tag{8}$$

4.2.1. Calculation of %QOS

This value, expressed as a % , measures how well the current network QOS values are satisfying the application QOS requirements. Since the specifications are measured in different units a non-dimensional representation like %QOS is needed to allow them to be in the same vector. This also has the benefit of scaling between zero and one. The requirements or specification values used for this research are given in Table 2: Application QOS Specification.

Table 2: Application QOS Specification

App	Thruput (bps)	Delay (msec)	Jitter (msec)	Loss Ratio (#dropped/total in period)
DB (query .75)	50kbps	<400ms	n/a	<1%
DB (entry .25))	n/a	<400ms	n/a	<1%
email (recv batch of 5)	n/a	<1000ms	n/a	<5%
email (send batch of 3)	n/a	<1000ms	n/a	<5%
Web Search	50kbps	<1000ms	n/a	<1%
Web images	500kbps	<1000ms	n/a	<1%
FTP (50% get 50% put)	n/a	<1000ms	n/a	<5%
Printer (B/W images)				
VOIP (G729)	100kbps (for 4 calls)	<15ms / node (10 hops)	25ms	<1%
Vid Conf Interactive video	500kbps	<15ms / node (10 hops)	30ms	<1%

The existing QOS metrics are most commonly "good" when low, e.g. Packet Loss, Delay, Delay Variance (Jitter), yet a few, very significant values are "good" when higher, e.g. throughput, upload speed, download speed. This necessitates different formulations when normalizing and scaling the data. Section 3.2 discusses these issues.

Once each QOS metric has been calculated the %QOS vector for time (*t*) is used to calculate an overall score between 0.0 and 1.0 to determine if the end-user is having a "Good" experience or a "Poor" experience based on the behavior of that node. This is necessary because network operations perspective on performance management is to monitor links and nodes which is distinct from applications developers who judge performance on an end-to-end basis cumulatively. Development of the QOEapp qualitative index will need future research but for this first approximation some values calculated based on theoretically Good or Poor experiences were used (see Figure 30) which must be validated by the human factors community.

QOEnode takes a vector of the QOEapp scores for time (*t*) to represent the current mix of traffic. The QOEnode uses a power value (*q*) to in some way quality the importance of that node to the overall service level agreements that have been established. For this first approximation powers were assigned based on how many end-users depended on that node (q=0.5 highest service level, q=2.0 average service level, q= 5.0 tolerable service level). In a real-word environment it would be up to Network Management to establish their preferred basis for selecting the power value and establishing a qualitative index to map the QOEnode scores to satisfactory, OK, unsatisfactory. For this study the same index as QOEapp is used.

4.2.2. Evaluation of QOEapp and QOEnode

These metrics express a qualitative performance value derived from the quantitative QOS values. Preliminary investigation of the formulation was in Excel® using some ‘known good’ to ‘known poor’ values as a proxy for actual QOS values that would be generated from the test-bed or live network. ‘Known good’ is used here to indicate that all application requirements for QOS values were met.

Analysis of the relationships indicated some modifications the lp-norm equation were needed to maintain the real-world behavior desired. As a simplification for this first approximation the QOS values (throughput, delay, jitter, packet loss) were all weighted equally at 1.0 and all used the same power appropriate to the established tolerance of this application for poor performance, e.g. intolerant VOIP users p= 0.2, HTTP = 2.0, very tolerant FTP users p=5.0

.

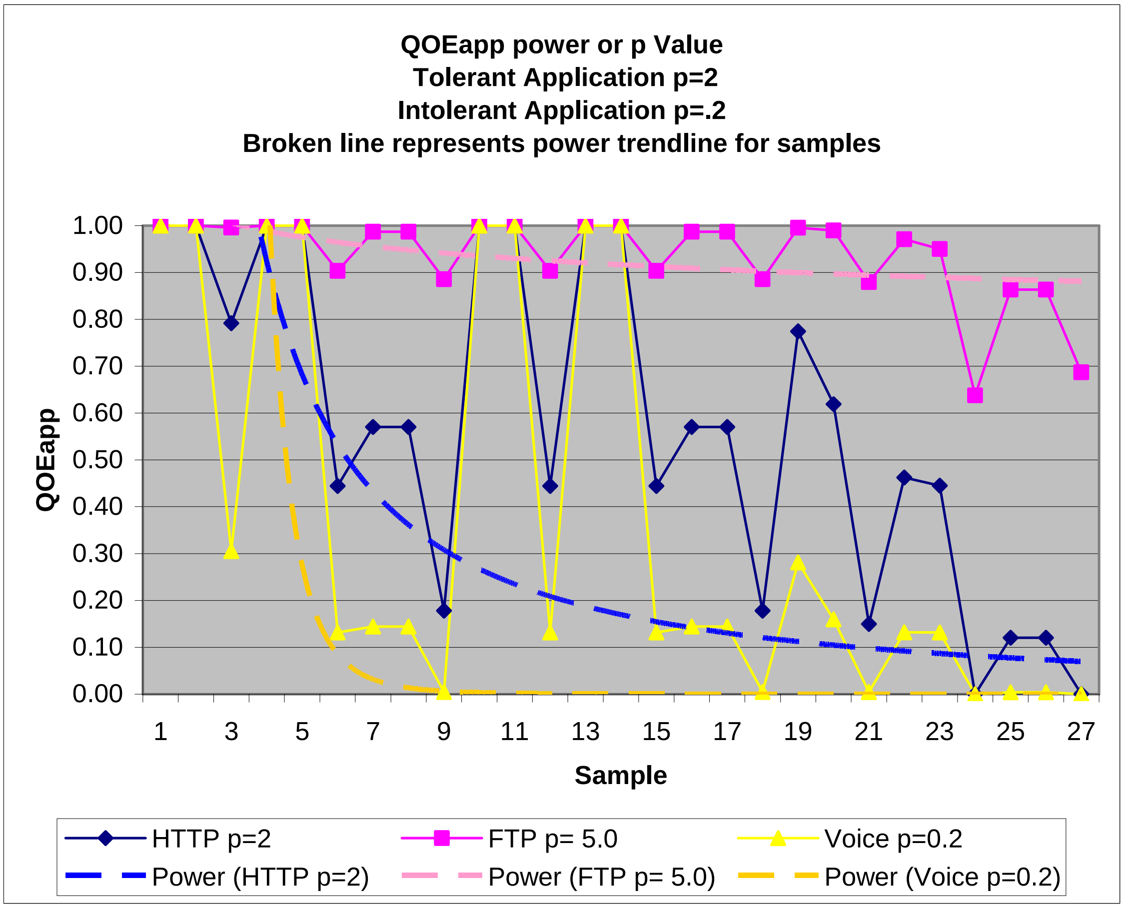


Figure 29 Data to establish the power curves

To establish the performance index (good to poor performance) the QOS values were examined and classified according to acceptable/unacceptable limits reported in the literature[74] see Figure 30. These are all first approximations and a separate investigation will be needed to finalize a rigorous application performance index.

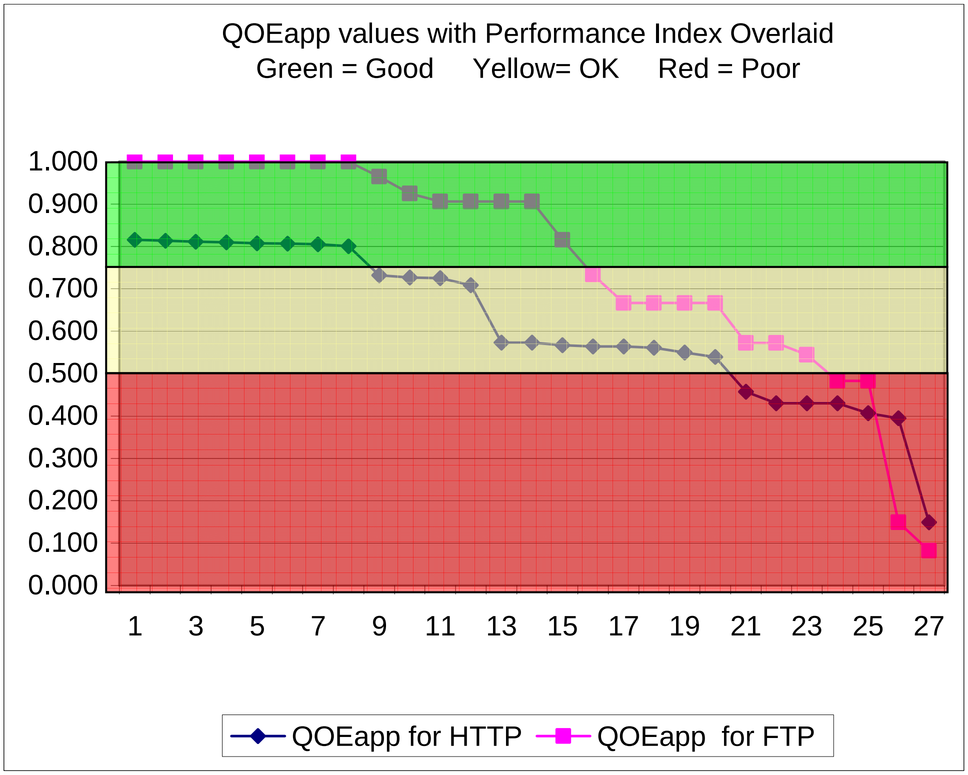


Figure 30: QOEapp values classified Good to Poor

Similarly there is future work with the weighting factors that could give one QOS value more significance than another in a particular application, For example if some application designer adds extra buffering to avoid packet loss they might want to weigh the delay variable more heavily. All this implies communication between application developers and network operations management but even before that occurs equipment providers need to collect at the node and provide the application specific QOS data rather than the current calculations based on undifferentiated aggregates of all bits. In the Cisco Network Analysis Module (NAM) some aspects of this approach are beginning to be implemented.

4.3. Utilization of an Application-Centric Key Performance Indicator

Once a credible test-bed is available well characterized applications can be run and the independent variables modified to create experimental scenarios. The simulation collects typical statistics for link, nodes and application, these are the dependent variables. A basic template will be used for this discussion

1. description of the scenario
2. simulation results from 3 perspectives:
   2.1. QOS statistics that network operations monitor (link capacity utilization)
   2.2. QOS statistic that application developers monitor (end-to-end delay)
   2.3. QOEapp value (qualitative measure of good to poor performance)
3. discussion of the results

4.3.1. Scenario 1: Baseline

The focus for these preliminary experiments is one switch in one network serving one building. The basic topology and link load have been verified and validated as described in 4.1 Additional LANs were added as in Figure 31. The simulation runs are set for15 minutes (900 sec) simulated network time, taking approximately 30 minutes clock time to complete.

There are 100 workstations in 4 LANs stochastically generating FTP, HTTP (both web browsing and image browsing), email, printing, and data base transaction discrete event packet level traffic. All this traffic goes to the Server linked to switch A_3548_209C_Slave using the TCP/IP protocol. The magenta arrow is the path from one LAN to the Server. There are also 5 Video Conferencing workstations and 5 VOIP workstations all starting UDP/IP sessions according to an exponential distribution.

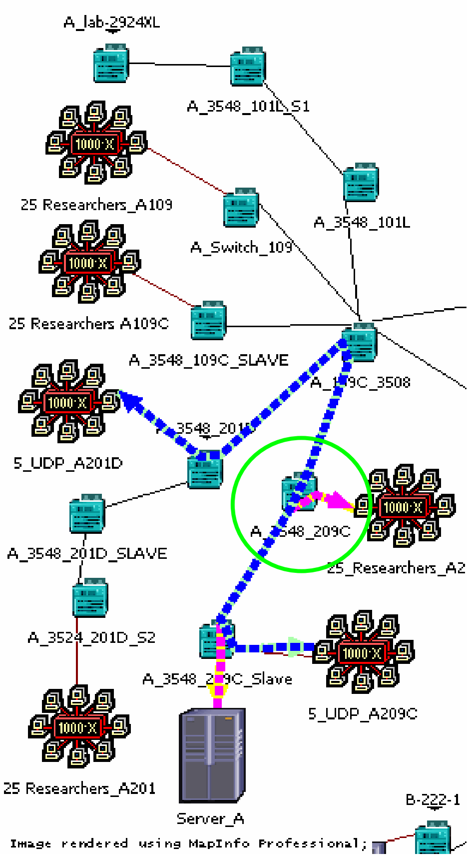


Figure 31 Focus of 1st Series of Experiments

4.3.1.1. Results: Baseline Over-Provisioned Scenario

Network operations, whether centralized or dynamically distributed in intelligent agents, will monitor only one or two key statistics. Capacity utilization is one and the threshold for problems is considered to be >20% over an extended period and growing. Figure 32 shows that in the over-provisioned scenario even with 2 video conference sessions it is less than 3% of the 1Gigabit link.

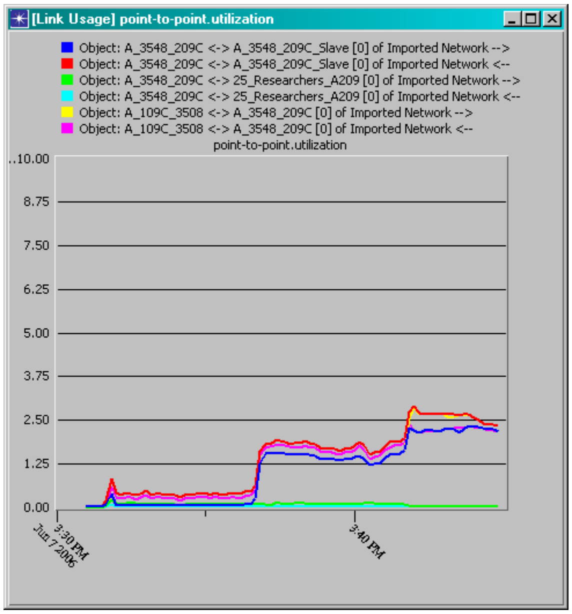


Figure 32: Link Capacity Utilization

From the application providers perspective the even the large image transfers are arriving well with in requirements, <100msec. Examining the queuing delay at the ingress and egress ports there is no significant delay.

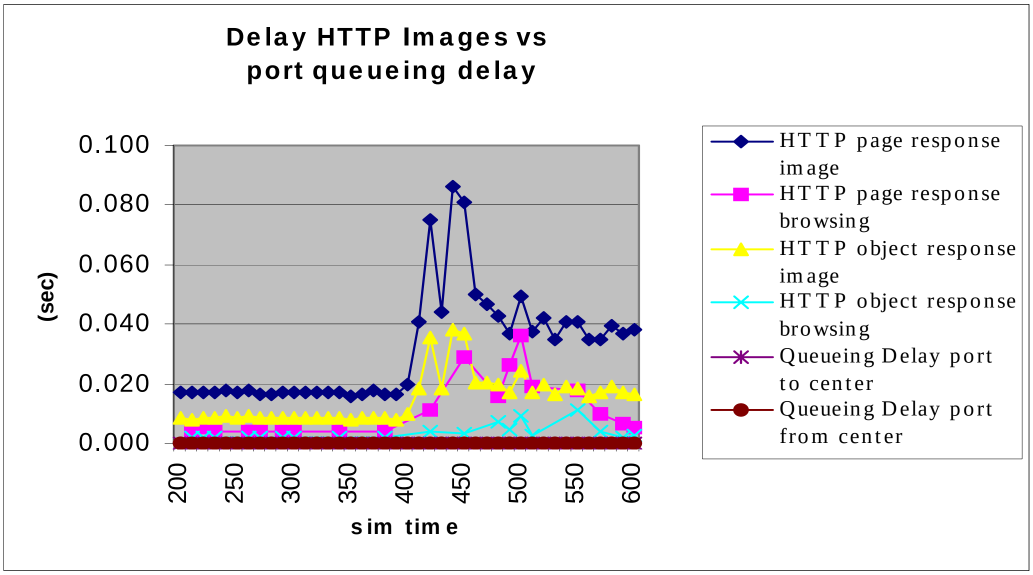


Figure 33 Web Application with <100ms delay

As representative applications just the HTTP and VOIP QOE values are calculated. QOEapp for each is at the maximum quality of user experience. This is consistent with the network operations and the application developer’s indicators.

Baseline Over-provisioned	**QOE HTTP**	**=**	**1**
Average 15 min	**QOE VOIP**	**=**	**1**

8.62E-06	Port Delay in/out HTTP
7.43E-06	Port Delay in/out Vid_Voip
9.9E+08	Available Thruput HTTP to Server
9.9E+08	Available Thruput Vid_Voip LAN 201 to 209
9.87E+08	Available Thruput HTTP from Server
9.87E+08	Available Thruput Vid_Voip LAN 209 to 201
0	Packet Loss Ratio HTTP
0	Packet Loss Ratio Vid_Voip
0	
-2.62E-10	Voice Jitter

Figure 34 QOE for Over-provisioned Network

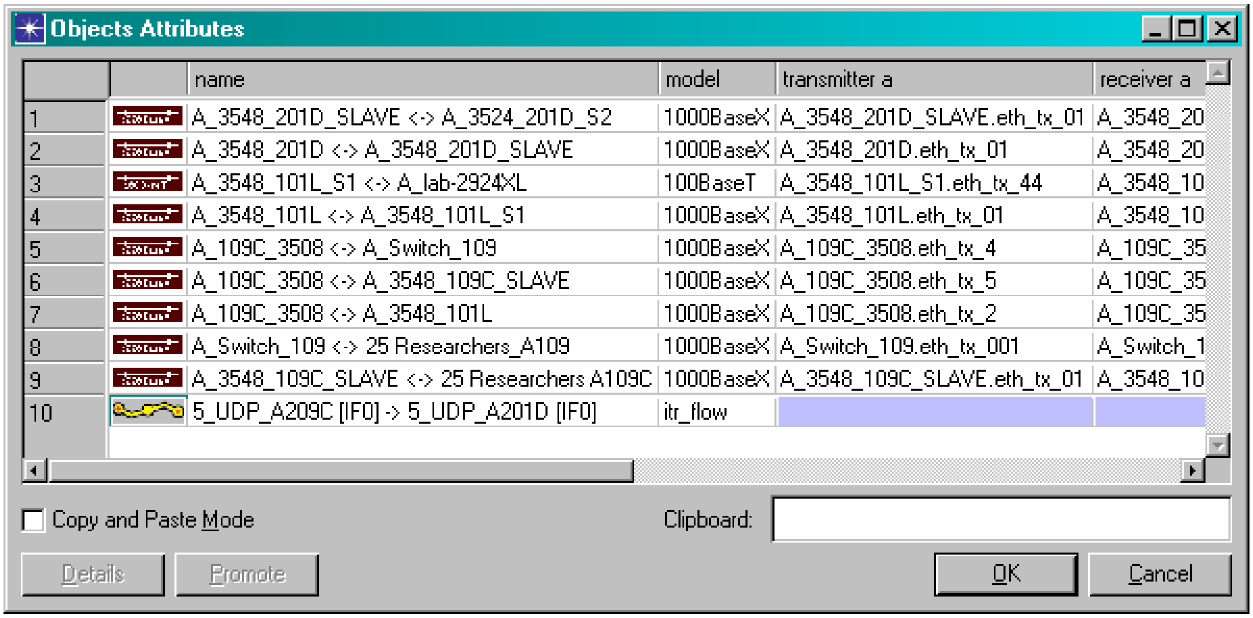
4.3.1.2. Discussion: Baseline Over-provisioned Scenario

There are no surprises here. Current indicators and proposed QOE performance indicators are in total agreement.

4.3.2. Scenario 2: Under-provisioned Links

Under-provisioning all the links in Network A models two possible real-world situations. It could be that the links were not upgraded due to economic considerations or that the traffic increases to utilized more capacity.

In the simulation it is simply a matter of changing the link model used for Network A. In Figure 35 all links are gigabit Ethernet except for line 3 which has already been reduced by an order of magnitude to 100Mbit or Fast Ethernet.



		name	model	transmitter a	receiver a
1		A_3548_201D_SLAVE <-> A_3524_201D_S2	1000BaseX	A_3548_201D_SLAVE.eth_tx_01	A_3548_20
2		A_3548_201D <-> A_3548_201D_SLAVE	1000BaseX	A_3548_201D.eth_tx_01	A_3548_20
3		A_3548_101L_S1 <-> A_lab-2924XL	100BaseT	A_3548_101L_S1.eth_tx_44	A_3548_10
4		A_3548_101L <-> A_3548_101L_S1	1000BaseX	A_3548_101L.eth_tx_01	A_3548_10
5		A_109C_3508 <-> A_Switch_109	1000BaseX	A_109C_3508.eth_tx_4	A_109C_35
6		A_109C_3508 <-> A_3548_109C_SLAVE	1000BaseX	A_109C_3508.eth_tx_5	A_109C_35
7		A_109C_3508 <-> A_3548_101L	1000BaseX	A_109C_3508.eth_tx_2	A_109C_35
8		A_Switch_109 <-> 25 Researchers_A109	1000BaseX	A_Switch_109.eth_tx_001	A_Switch_1
9		A_3548_109C_SLAVE <-> 25 Researchers A109C	1000BaseX	A_3548_109C_SLAVE.eth_tx_01	A_3548_10
10		5_UDP_A209C [IF0] -> 5_UDP_A201D [IF0]	itr_flow		

Figure 35 Under-provisioning the Links by an Order of magnitude

Other then that one modification topology and the number of end-users remains the same as in the Baseline.

4.3.2.3. Results: Under-provisioned Links Scenario

The 20% threshold has been crossed and if this persists or a customer complains it could cause network operations to attempt to diagnose a problem. With 3 video conferencing sessions, out of a potential 10, started during this 15 minute interval, utilization reaches 40% of capacity. Some of these will terminate during the next interval and others will start.

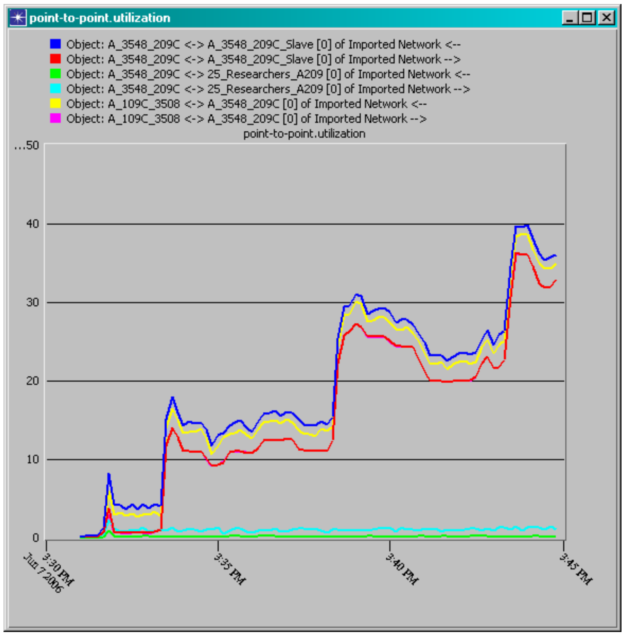


Figure 36: Utilization rises above 20% after 2nd Video Conference begins

From an application developer view point the end to end performance of web browsing has improved and is no cause for concern. The reason for this behavior not clear to this researcher.

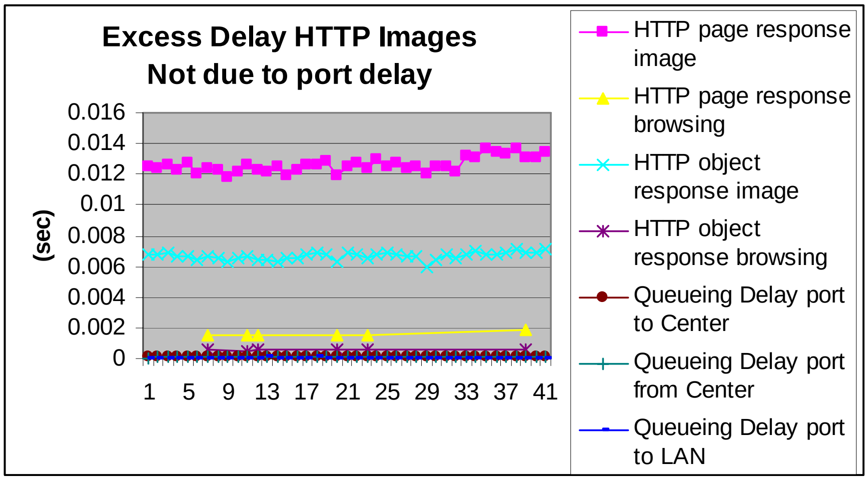


Figure 37: Delay is considerably lower than average in Baseline

It is unlikely t be a stochastic anomaly due to using the mean of multiple runs. It is more likely to be normal, but a network teletraffic engineer or application expert would have to determine that. On the other hand it could be unrealistic and a flaw in the OPNET models.

Link Reduction 1			QOE HTTP =		1
Average 15 min			QOE VOIP =		1
6.18E-05		Port Delay in/out HTTP			
1.28E-04		Port Delay in/out Vid_Voip			
81,703,374		Available Thruput HTTP to Server			
81,697,966		Available Thruput Vid_Voip LAN 201 to 209			
78,163,035		Available Thruput HTTP from Server			
78,163,035		Available Thruput Vid_Voip LAN 209 to 201			
0		Packet Loss Ratio HTTP			
0		Packet Loss Ratio Vid_Voip			
1.20937E-08		Voice Jitter			

Figure 38 QOE for Under-provisioned Links

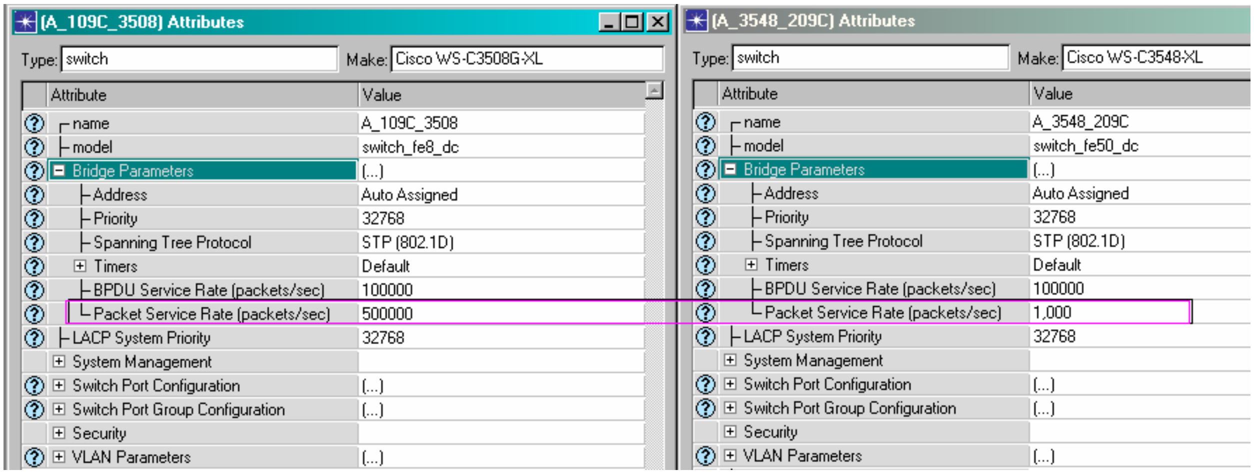
Although the delay, throughput and jitter have increased over the baseline scenario they don’t exceed the limits established by the application engineers and listed in Table 2: Application QOS Specification). Consequently the QOEapp values show good performance.

4.3.2.4. Discussion: Under-provisioned Links Scenario

In this scenario the network operations indicator shows a more “serious” performance impact than either application or QOEapp indicators. When comparing the QOE results of under-provisioned links (Figure 38) to over-provisioning (Figure 34) there is an impact. This suggests that another experimental scenario must be designed to degrade performance further. Rather than reduce the link capacity another order of magnitude to 10xT, which would be unrealistic, more workstations will be added to the LANs to increase traffic.

4.3.3. Scenario 3: Under-provisioned Nodes

To produce a scenario where an under-powered switch, from either old equipment or damaged CPU, degrades performance the switch attribute for packet service rate was reduced to servicing only 1000 packets/sec. This was accomplished by modifying the switch attribute “packet service rate” as shown in Figure 39



(A_109C_3508) Attributes

Type: switch Make: Cisco WS-C3508G-XL

Attribute	Value
name	A_109C_3508
model	switch_fe8_dc
Bridge Parameters	(...)
Address	Auto Assigned
Priority	32768
Spanning Tree Protocol	STP (802.1D)
Timers	Default
BPDU Service Rate (packets/sec)	100000
Packet Service Rate (packets/sec)	500000
LACP System Priority	32768
System Management	
Switch Port Configuration	(...)
Switch Port Group Configuration	(...)
Security	
VLAN Parameters	(...)

(A_3548_209C) Attributes

Type: switch Make: Cisco WS-C3548-XL

Attribute	Value
name	A_3548_209C
model	switch_fe50_dc
Bridge Parameters	(...)
Address	Auto Assigned
Priority	32768
Spanning Tree Protocol	STP (802.1D)
Timers	Default
BPDU Service Rate (packets/sec)	100000
Packet Service Rate (packets/sec)	1,000
LACP System Priority	32768
System Management	
Switch Port Configuration	(...)
Switch Port Group Configuration	(...)
Security	
VLAN Parameters	(...)

Figure 39 Packet Service Rate configured rate on Left to rate on Right

The value of 1000 was selected based on a review of the packet/sec sent from the LANs. to the switch with reduced service rate. Interestingly, when the applications are evaluated for bits/sec throughput Video Conferencing dominates due to the large packet size. In Figure 40 Application traffic from LANs packets / sec it is clear the VOIP dominates if the packets/sec is considered.

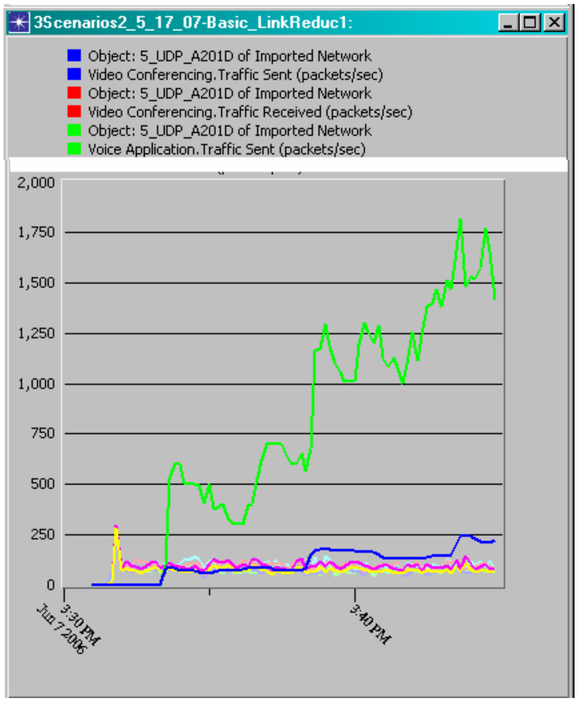


Figure 40 Application traffic from LANs packets / sec

4.3.3.5. <u>Results: Under-provisioned Nodes Scenario</u>

There is nothing in the link utilization statistics to cause network operation to investigate the switch for poor performance. In fact 1.25% utilization is even lower than the over provisioned scenario 1.

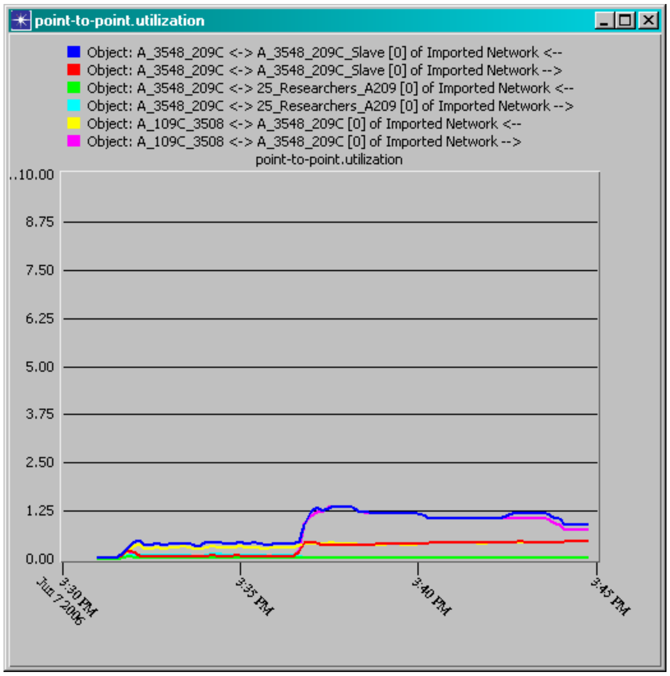


Figure 41 Scenario 3 Capacity Utilization

However, the HTTP application users will be complaining soon since the page response starts climbing at sample 20. This is seen in Figure 42

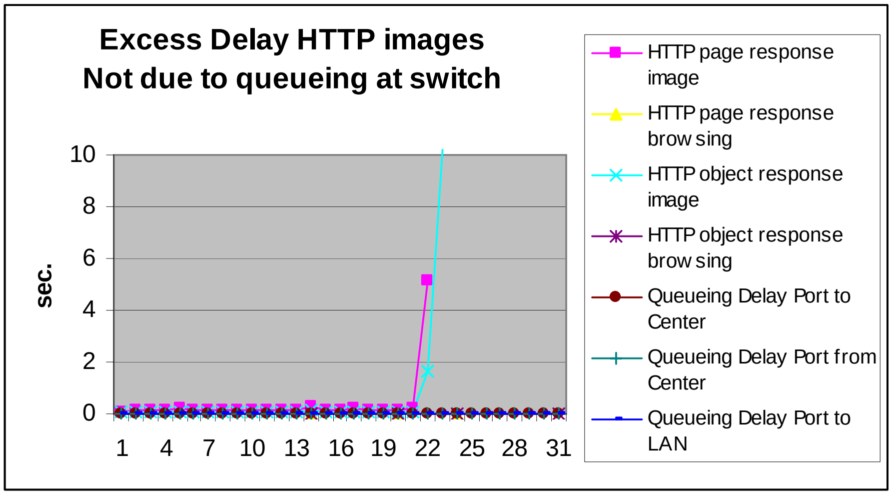


Figure 42 Web browsing response time very high

For the application engineer there are some reasons for concern. The large HTTP image objects are being delayed aver 10 seconds when competing for switch resources with VOIP and

Video Conferencing. Since they are UDP applications with no retransmission scheme, switches typically prioritize UDP packets over TCP application packets.

Node Reduction 1			QOE HTTP =		1
Average 15 min			QOE VOIP =		1
9.78E-06		Port Delay in/out HTTP			
9.23E-06		Port Delay in/out Vid_Voip			
988,731,491		Available Thruput HTTP to Server			
988,725,542		Available Thruput Vid_Voip LAN 201 to 209			
985,240,054		Available Thruput HTTP from Server			
985,240,054		Available Thruput Vid_Voip LAN 209 to 201			
0		Packet Loss Ratio HTTP			
0		Packet Loss Ratio Vid_Voip			
-1.29E-09		Voice Jitter			

Figure 43 QOE for UDP vs. TCP service resources scenario

QOEapp does not show a degradation of performance for HTTP which would seem to be occurring. Port delay is higher for HTTP than VOIP yet it still isn’t very high at a single node.

4.3.3.6. Discussion: Under-provisioned Node Scenario

QOEnode was never explicitly calculated since all QOEapp values indicate all applications performing well. It is given that if all QOEapp are good QOEnode will be satisfactory rather than unsatisfactory. This preliminary experiment raises the question of what statistic would inform QOEnode that UDP and TCP are contending for resources in an unsatisfactory manner for web users waiting over 10 sec for images?

4.3.4. Forcing a Failure Scenario

The three simulation scenarios while portraying real-world conditions did not exercise the proposed QOE indicators. As discussed above this may be a limitation of the models or of the statistics calculation which can be remedied over time. On the other hand it may be a well modeled simulation scenario which is demonstrating that the QOE indicators will highlight no switching decay in performance when TCP/IP is implemented at the client and server end-nodes. This would be unfortunate since the root causes of poor performance are under-provisioned links and nodes which is within the Network Management domain of responsibility not the end-users.

Evaluating the QOE KPI in simulation will require an extensive series of experiments which need some justification before they are undertaken. While investigating the lack of switch and router reaction to under-provisioning the values from the over-provisioned (i.e. current campus configuration) simulation can be manipulated in Excel to evaluate the QOE indictors supposing conditions can be modeled that will degrade the QOS metrics sequentially. Again there are endless permutations of this type of mathematical testing. Rather than a mathematical demonstration of utility, the use of real-world scenarios and demonstrating the utility of QOE under realistic conditions is preferred. The resource demands of the different levels of traffic are not negligible.

Figure 44 demonstrates the reactions of QOEapp to steady degradation of QOS performance metrics. Delay is steadily increased by 11msec, Packet Loss by .05, Jitter by .5msec, and bandwidth is decreased in increments of 10Mbps.

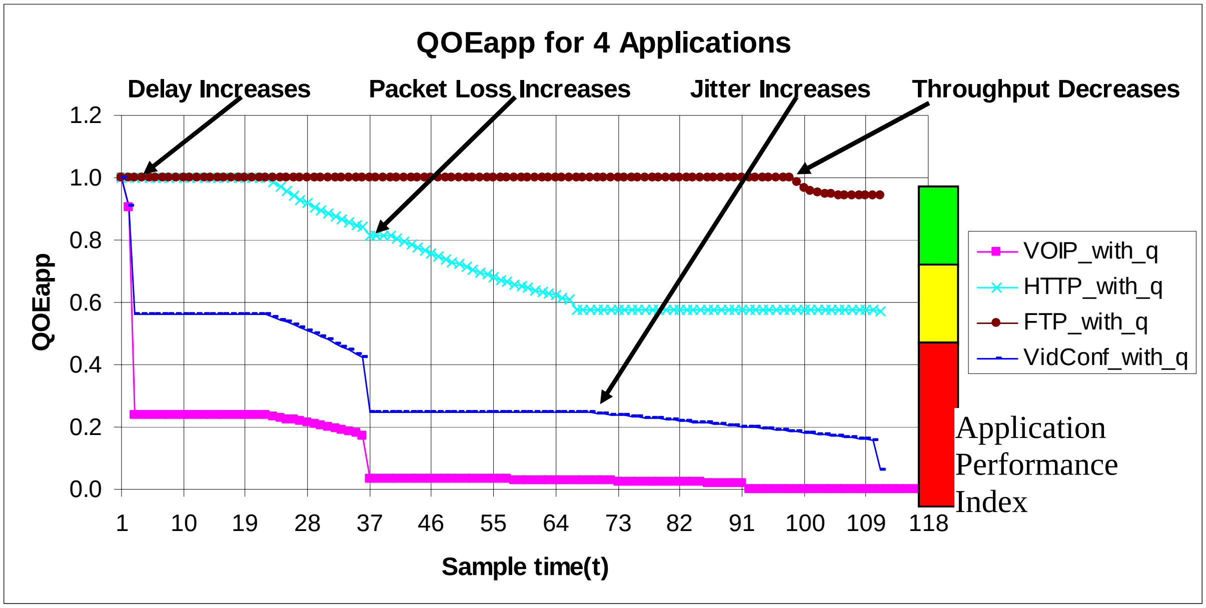


Figure 44 QOEapp reaction to degrading Performance

As anticipated an application such as FTP is insensitive to all performance degradation except reduced bandwidth and stays in the "Good" performance range. HTTP QOE degrades for an "OK" range when packet loss and delay are present and stays there until throughput degrades. Video Conferencing is sensitive to delay yet with buffering can maintain an OK level of experience until the Packet Loss degrades as well (around sample 35). Finally a VOIP user would experience "Poor" quality as soon as the Delay exceeds 33msec. However the application QOE is significant when diagnosing the cause of QOEnode "Unsatisfactory" performance.

The three simulation scenarios did not provide values that would test the QOEnode formulation but using the forced degradation spreadsheet it can be calculated. Each sample *t* represents the mix of application traffic at a moment in time and the QOEapp values for that time period, consequently a QOEnode can be calculated. Initially, samples representing various performance experiences for the end-users will be used and the value of "q", the power element, is varied based on the number of users serviced by the node. If the node services many users it would have a “High” service level agreement (SLA) if only a few then a “Low” SLA. In reality Network Management may have some other criteria to assign significance to the node.

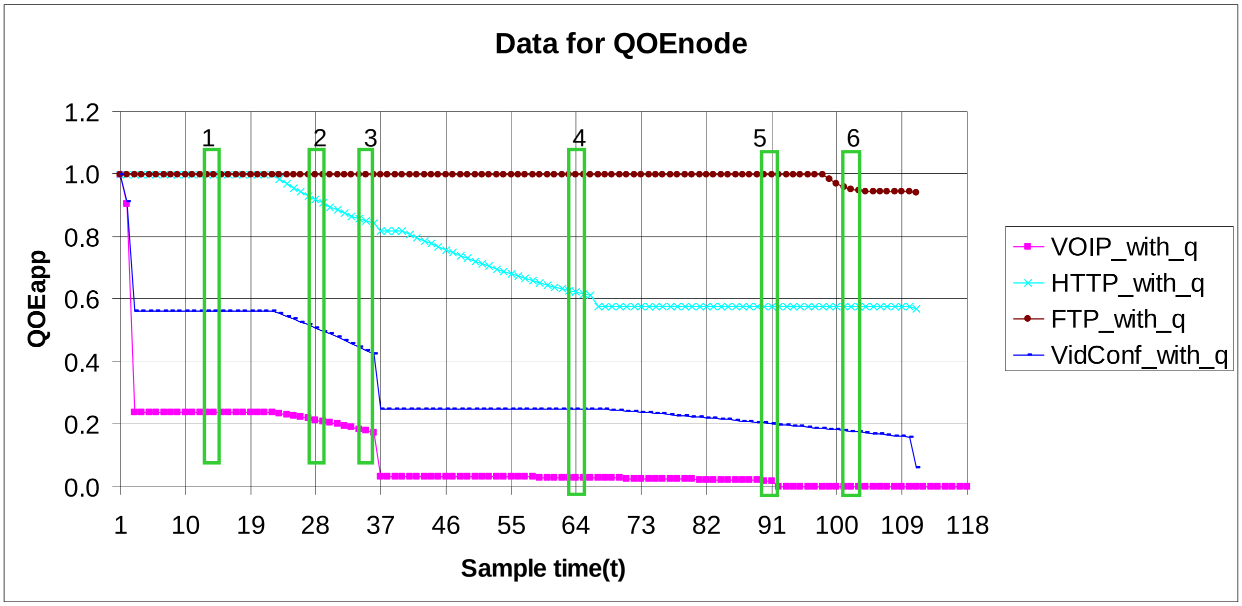
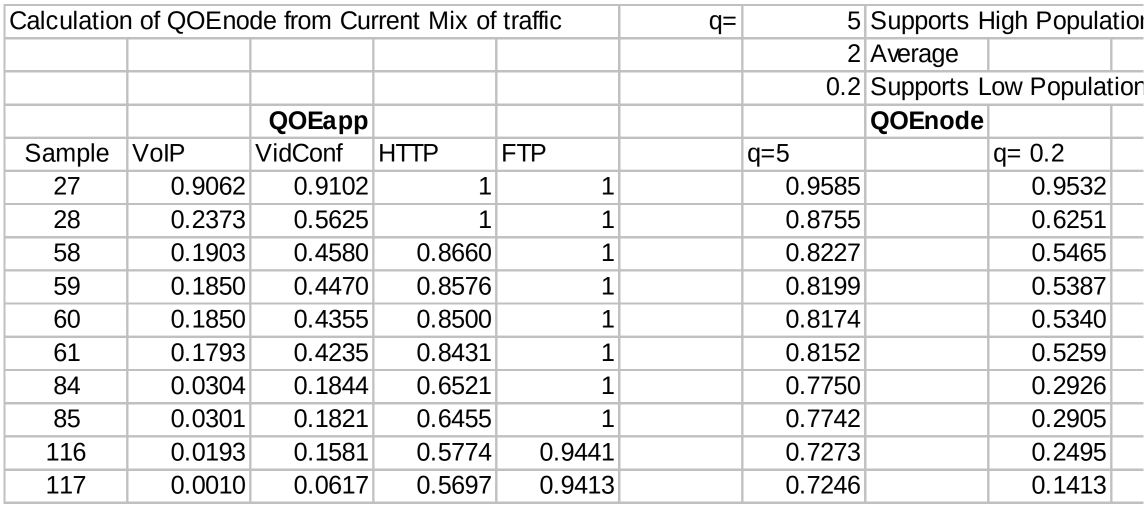


Figure 45 Twelve samples selected for QOEnode Calculation

Two or more samples were selected from the inflection points as shown in Figure 45 and used in the lp-norm-like formula for QOEnode.

$$\mathrm{QOEnode} = \left(\left(\sum_{i-n} \left| \mathrm{QOEapp}_{i-n} \right|^{q} \right) \Big/ n \right)^{1/q} \qquad (9)$$



Calculation of QOEnode from Current Mix of traffic					q=	5	Supports High Populatio	
						2	Average	
						0.2	Supports Low Population	
		QOEapp					**QOEnode**	
Sample	VoIP	VidConf	HTTP	FTP		q=5		q= 0.2
27	0.9062	0.9102	1	1		0.9585		0.9532
28	0.2373	0.5625	1	1		0.8755		0.6251
58	0.1903	0.4580	0.8660	1		0.8227		0.5465
59	0.1850	0.4470	0.8576	1		0.8199		0.5387
60	0.1850	0.4355	0.8500	1		0.8174		0.5340
61	0.1793	0.4235	0.8431	1		0.8152		0.5259
84	0.0304	0.1844	0.6521	1		0.7750		0.2926
85	0.0301	0.1821	0.6455	1		0.7742		0.2905
116	0.0193	0.1581	0.5774	0.9441		0.7273		0.2495
117	0.0010	0.0617	0.5697	0.9413		0.7246		0.1413

Figure 46 QOEapp used to calculate QOEnode

The Node performance index is for now the same as for QOEapp although it would normally be based entirely on the network management criteria for a particular access network. The power values are reversed for QOEapp and QOEnode. For an application a .2 is selected for an application where a good user experience is desired. For a switch or router a .2 in this scheme represents a node with few connections and consequently a low priority for performance monitoring.

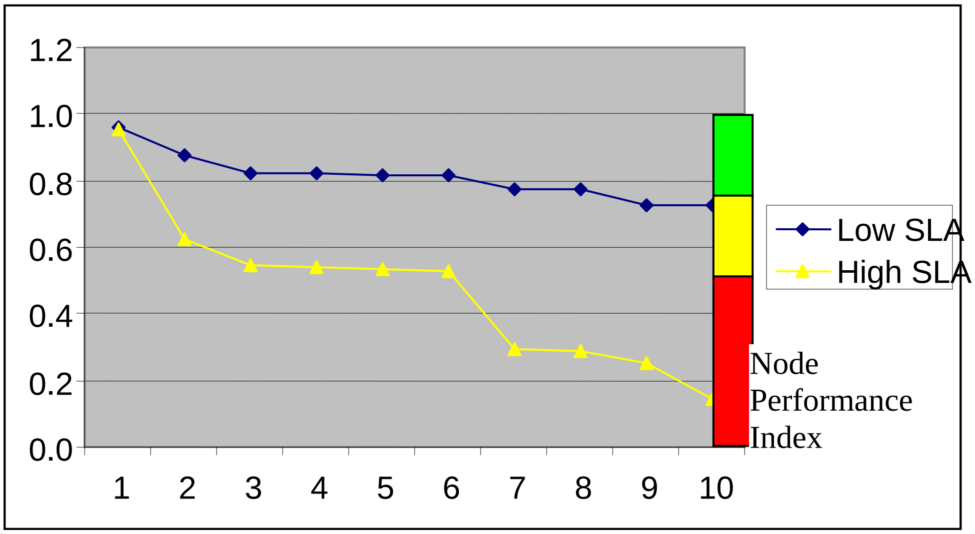


Figure 47 The same traffic mix with only the SLA power is changed

4.3.5. Summary of Preliminary QOE Utility investigation

QOEnode was never explicitly calculated for the simulations since all QOEapp values indicated all applications performing well. It is given that if all QOEapp are good QOEnode will be satisfactory rather than unsatisfactory. In Excel the QOEapp values could be forced to degrade by increasing the switch value for packet loss. It appears as though the switch model does not calculate dropped packets.

From the 3 scenarios it can be seen that QOE indicators always agree with one of the other 2 indicators, i.e. network operations which monitors the more device-centric link capacity utilization or the application's end-to-end response time or delay. The intent of a QOE indicator at the node level is to give network performance management a switch by switch monitor to detect developing customer end-to-end performance dissatisfaction and pin point where in the network the degradation is located. It is encouraging that QOE could serve as a tie breaker between the network and application indicators. However, the critical link that would tie QOE to end-user dissatisfaction will require multidisciplinary research with HCI and HF personnel. Similarly, working with network operations collection of these QOS application specified statistics may need to be augmented with other more device oriented statistics like a new %UDP packets value.

Finally the Cisco switch model offered by OPNET does not permit reduction of queue size at the media access connection (MAC) level for a single port. The MAC object is global and reducing the queue size from infinite to 1000 packets on one port modifies all ports in the simulation. The object hierarchy can be changed and the module recompiled but that effort was not undertaken at this time. In the forced failure mode, the output from reducing the node processing, showing no packet loss due to infinite queue size, was modified in the spreadsheet to show gradually increasing packet loss until the node was clearly unable to support “Good” or “OK” application performance.

The forced failure of applications then gives an interesting variety of values for applications and the QOE can be calculated for that node to indicate satisfactory/unsatisfactory performance against some performance index and calculated according to a given level of service power function for the criticality of that switch.

CHAPTER 5: CONCLUSIONS AND FUTURE WORK

Do not go where the path may lead, go instead where there is no path and leave a trail
Ralph Waldo Emerson

The fundamental conundrum addressed in this dissertation is the gap between modern telecommunications usage, which is pervasive and application-centric in nature, and the device-centric network metrics used by network operations to maintain the telecommunications infrastructure and customer's satisfaction. The prevailing school of thought is that more bandwidth and more intelligent protocols will address all performance issues. Even if this is essentially true, with the constant changes in traffic and topology network management would still need to measure and evaluate current performance. In the eyes of this researcher, the area is a complex socio-technical system currently shifting from a technical infrastructure focus to a more customer, i.e. application, policy based view. This dissertation takes a broad sweep through this problem space evaluating an approach for new indicators and simulation tools available to investigate a shift from device-centric metrics to application-centric metrics in response to end-user's quality requirements. Contributions touch the areas of:

- Network Management which may become more proactive by forging a tighter link between customer application requirements and network Service Level Agreements. Since computer science research in distributed applications is improving the ability of applications to deal with network congestion, delay, etc. a competitive advantage may exist for network management organizations which shift to focus on the performance of specific mission critical application and customer quality of experience (QOE).

- New application-centric indicators to support this management view of the network, where infrastructure equipment or appliances might calculate the two proposed node statistics as triggers for performance problem diagnosis. The first, QOE_{node}, captures the subjective customer satisfaction with the current mix of application traffic flowing though the node. The second, QOE_{app}, evaluates each application class as a contributor to QOE_{node}. This may also be a valuable metric when troubleshooting poor performance.
- The proposed QOE metrics are based on the current device-centric metrics and inspired by uses of lp-norm and functional analysis. Independent variables have been identified that can be used to run controlled experiments. The variable “provisioning” was varied to conduct a few initial simulations to test the utility of the QOE indicators.
- Recently the validity of the network research conducted using simulations has been questioned consequently the development of a credible simulation test-bed is documented and future research conducted in this or a similar test-bed will add to research credibility.
- For network researchers, either academic or industrial, there is always a challenge when gathering empirical data due to concerns about network security. Network Management is understandably reluctant to allow collection of actual network topology and traffic. A procedure was worked out with our IT group which created the trust needed.

Team contributions/solutions are needed to cover all the dimensions of this problem: device vendors, and researchers in network management with both a business (policy) and infrastructure view, application architecture and design computer scientists, mathematicians, teletraffic engineers, and human computer interactive systems engineers.

5.1. Conclusions

As mentioned in the first chapter, this research is a response to the NRC 2001 call for a focus on network operations research:

> Over the past three decades, several bodies of theory, such as performance analysis and resource allocation/optimization, have contributed to the design and understanding of network architectures, … However, as the Internet has evolved into a critical infrastructure used daily by hundreds of millions of users, *operational concerns such as manageability*, reliability, robustness, and evolvability have supplanted the performance of the data forwarding plane as the limiting factors. … theoretical understanding of these crucial areas is poor, particularly in comparison with their importance. [5] *(emphasis by author)*

Creating a research laboratory with real-world equipment and applications is expensive. Consequently "real" test-beds tend to be small. To be representative of today's telecommunications networks simulations are the typical research environment. This can be created from theoretical configurations of nodes and traffic or from empirical data. The pitfall of theoretically generated networks is that the results may be purely descriptive of "some way" to generate the statistical behavior of a network but which may have no intrinsic relationship to the physical constraints found in live networks. Empirically-based research simulation test-beds must including the topology, specific application traffic, and general network traffic accurately modeling the statistical behavior of real-world deployed networks. As with any test-bed, it must be calibrated and well understood by the researchers before experiments are run. For this research I created a moderately large scale network simulation using a widely available tool, OPNET. This test-bed was validated against the real network segments which it models, and in the process I developed the necessary procedural tools and methods required to extract data and to effectively work with network administrators in a real network. The simulation and procedures have been well documented and will be available and useful for future research.

I proposed, developed, analyzed, prototyped, evaluated, and documented new network management metrics that have the potential to be an important methodology for real-time management of congested networks. The simulation test-bed was developed importing actual switch and router configurations and link traffic that was validated by campus network management. Another form of contribution useful for future researchers were discoveries about the simulation environment, OPNET. In the process of verifying behaviors I located a bug in the simulation software in the modeling of Cisco Systems proprietary protocol Hot Standby Router Protocol (HSRP). There were several developer errors in configuration of client/server traffic resulting in unanticipated results. These discrepancies had to be resolved before running controlled experiments on over-provisioned /under-provisioned networks and the utility of the proposed QOE indicators. An important, often relearned lesson for researchers and developers alike: even very well designed and implemented software such as OPNET can contain latent bugs that can impact results or delay development.

An initial set of scenarios to evaluate the utility of the proposed QOE performance indicators gave promising but inconclusive results. This is clearly a first approximation of an L^p-norm-like approach since no effort was made to use weighting factors for the contribution of each QOS metrics to end-user experience. Similarly no attempt was made to weight the traffic mix to represent the proportion of the traffic from each application. Questions were raised that can only be answered in collaboration with network operations, application designers, mathematicians and human computer interaction experts. Such questions seem an inevitable result of research in novel domains.

5.2. Contributions

The contributions of this dissertation are:

- Postulated and demonstrated in simulations a shift in network management metrics from device-centric to application-centric:
    - It is feasible to proactively manage the network through a deeper understanding of customer satisfaction as application performance degrades from (good to poor) as monitored at each node
    - Provide a feasibility demonstration of a tighter linkage between application specifications and service level agreements to evaluate advantages
    - Demonstrated that monitoring an indicator of service level compliance for the current mix of traffic (satisfactory to unsatisfactory) can serve as a trigger for centralized or distributed operations management to investigate performance, reliability or security failure
- Developed three specific metrics that map device-centric metrics to application-centric indicators
    - %QOS -links device-centric metrics to application requirements and specification
    - QOE_{app} –necessary intermediate mapping of device performance to performance of a class of applications.
    - QOE_{node} –suitable for monitoring the level of customer satisfaction at each device
- Designed, developed, and implemented a simulation test-bed of sufficient size and complexity for telecommunication experiments to demonstrate useful properties captured in scenarios

  - Represents a class of access networks, single vendor ~1000 end-users
  - Developed a procedure to establish trust with the Network Security Group to allow the collection of real-world device configurations and traffic flow
  - Investigated a leading telecommunications modeling and simulation software for ability to accomplish the task including
    - Import of topology
    - Import of traffic
    - Creation of experimental scenarios with well understood behavior
    - Collection of statistics to properly evaluate new QOE statistics
      - Validity of standard statistics
      - Validity of standard application traffic generators
      - Ability to create specialized probes for non-standard statistics

- Conducted a case study to evaluate the utility of the first approximation for QOE indicators:
  - Outlined a series of simulation experiments to determine the utility of QOE metrics based on four independent variables
  - Conducted the first of this series of experiments
  - Discovered and documented limitations and software errors in a widely used network simulation tool.

5.3. Future Work

Future work will relate to enhancing the formulation of the QOE indicators and the credibility and complexity of the test-bed.

The QOE indicators, which are first approximations with the advantage of corresponding to network artifacts, may be found to exhibit more L^p-norm characteristics when evaluated by mathematicians with a deeper knowledge of functional analysis. In addition computer science researcher well versed in the direction of distributed applications can determine if the %QOS vector, which weights all QOS parameters equally, would benefit by weighting them differently. This could be in collaboration with a human computer interactions (HCI) researcher to determine if the assignment of applications to power curves in QOE_{app} accurately reflects user tolerance to degraded performance. These HCI researchers and human factors (HF) researchers are needed to validate the performance index (good to poor quality of experience) with human factors research. Similarly, the QOE_{node} performance index (satisfactory to unsatisfactory switch/router performance) need to be validated with network management and evaluated given other application mixes in the traffic presented. This could be developed into a more application-centric basis for SLAs rather than the current average QOS metrics.

Following the credibility criteria proposed by [58, 59, 60] the simulation test-bed could be enhanced in future experiments.

Currently application traffic is generated using standard distribution curves (exponential) found in the OPNET traffic generation models. Since traffic has been found to be bursty and not always poison distributed, empirically derived distributions may be significant. Currently this is not significant since short time periods (15 minutes to one hour) are used in the scenarios. These empirical distributions can be created with OPNET®ACE or using Modeler v12.5 which directly imports Cisco NAM data.

Having a suitable pseudo random number generator is essential to any complex simulation. To date the standard OS implementation of the Berkley PRNG seems sufficient.

However if future scenarios seem to be biased by repetition of the PRNG the OPNET® user group has a java based PRNG with a longer number stream.

In additional scenarios appropriate tuning of device parameters/ configuration will create network stress conditions for all the independent variables, i.e. different application mix, equipment configurations, and mixes of equipment, as well as, different examples of under-provisioned networks. In addition, an ongoing task for all new scenarios will be clear documentation of the model changes and configuration control of the scenarios so there can be reproducible data analysis between the scenarios

As the standard analysis practices that need to be performed on each scenario are better understood, standard Analysis Configuration templates will insure consistent results. In addition there may be new node statistics and the QOE statistics which must be implemented as OPNET® custom statistic to eliminate the export of data to Excel®. This programming task should be undertaken once the QOE formulation seems relatively complete.

APPENDIX A: QUALITY SERVICE ARCHITECTURES

Application Classes of Service are proposed by many standards bodies. For this research a taxonomy of Application Classes will be worked out from the various standards (IETF, ITU, W3C, et al) to include applications such as, Distributed Training Simulations, Video Gaming, Video on Demand, etc. Classes of Service as specified by standards bodies are evaluated against empirical QOE reports [31].

The classes of service proposed by the ITU-T for IP telecommunications as related to hybrid networks (voice & data) is used. Found in ITU-T Recommendation Y.1541.

Table 3 ITU-T Class of Service with Application Examples

QOS Parameters / **Class of Service**	**Description**	**Mean Delay Upper Bound**	**Delay Variance (Jitter)**	**Packet Loss Ratio**	**Application Examples**
Class 0	Real-Time, Highly Interactive, Jitter Sensitive	100 ms	< 50ms	$<10^{-3}$	VoIP, Video Teleconferencing, Distributed Simulations
Class 1	Real-Time, Interactive, Jitter Sensitive	400 ms	< 50ms	$<10^{-3}$	VoIP, Video Teleconferencing Video Games
Class 2	Highly Interactive, Transaction Data	100 ms	Unspecified	$<10^{-3}$	Web Browsing Data Base
Class 3	Interactive, Transaction Data	400 ms	Unspecified	$<10^{-3}$	Peer-to-Peer
Class 4	Low Loss	1 sec	Unspecified	$<10^{-3}$	FAX, Bulk Data, Streaming Video
Class 5	Unspecified	Unspecified	Unspecified	Unspecified	Default IP Networks

ATM supports five different classes of service:

- Constant bit rate (CBR) allows the desired bit rate to be set when the virtual circuit is established; it is used for services such as uncompressed voice and video;
- Variable bit rate–non-real time (VBR–NRT) allows statistical techniques to be used to optimize network throughput when the rate at which data is available varies;
- Variable bit rate–real time (VBR-RT) is intended for applications such as compressed speech and video, where data delivery must occur at regular intervals;
- Available bit rate (ABR) is used for non-time-critical operations such as bulk file transfers that can adjust their rate of input to use available network capacity; minimum acceptable rates can be specified to ensure some service at all times.
- Unspecified bit rate (UBR) is the residual class with no guaranteed properties; it is used primarily for TCP/IP data traffic

APPENDIX B: TELECOMMUNICATIONS SIMULATION TOOLS

Table 4 Network Simulation Tools Considered

Tool Name	**Source**	**Strengths**	**Weaknesses**	**URL (July 2007)**
Research Tools				
Vint-NS2-Nam	USC / UCB / LBNL Xerox PARC	Robust; Commonly used in academia Public Domain	No Cisco models; Limited traffic models	http://www.isi.edu/nsnam/vint/
SSF	S3 (DARPA project) Renesys (commercial)	Parallel processing; Scales to >100k nodes Public Domain	Learning curve; Support; Robustness	http://www.ssfnet.org/
PADS/ various	Ga. Tech. Parallel and Distributed Simulations Dept.	Parallel processing; Academic Public domain	Learning curve Support Robustness	http://www.cc.gatech.edu/computing/pads/app-telecom.html
Commercial Tools				
OPNET	OPNET Technologies formerly MIL3	Free to University Researchers Ability to Import Configurations & MRTG Traffic Cisco Models Founded 1986	Complex OO Environment, long learning curve but free training in MD	http://www.opnet.com
MLDesigner	MLDesign Technologies	Models Systems from SOC to Networks Descendant of BONeS	Inability to import real-world data Less generous University support Founded 2000	http://www.mldesigner.com

APPENDIX C: NETWORK RESEARCH AUTHORIZATION

Policy and Procedure for Data Release and Use

Preamble:

Network researchers frequently find the lack of real world data an obstacle to advancing or testing research hypotheses. UCF Information Technologies and Resources will review on a case by case basis all requests from researchers and will accommodate their needs within the bounds of UCF IT security.

Purpose:

This policy addresses when, to whom, and in what form the Network Operations Center of Computer Services releases network data and also outlines appropriate uses of data and release procedures.

Scope:

Includes the resources under the management or control of Computer Services/Network Operations

Policy:

Data users must inform Network Operations Center (NOC) who are responsible for collecting the data, and confirm with NOC that their research plans are not in conflict with existing NOC research and publications

Data users must agree not to distribute the data to others without prior written authorization

Data users must agree to give proper acknowledgement to the NOC of Computer Services as well as the data collection team

Data users must absolve the Network Operations and Computer Services and the data collection team of any responsibility for inadvertent errors in the data that lead to wrongful analysis and improper decision making

Data users must agree to keep the data on a secure system using firewalls, antivirus software and unique ID for access to the system and data

Data users must agree to destroy the raw data once the research is complete

Data Collection Team must make sure that the data released does not contain sensitive data elements that would violate state or federal laws; e.g. FERPA, HIPAA, GLB, etc.

Data Collection Team must inform the Data user of the nature of sensitivity associated with the data to be released

Data users must agree not to publish data classified as “Moderately Sensitive” or higher on the “Data Classification Standard” document; e.g., IT infrastructure diagrams, network diagrams, core IP addresses, etc.

Data users must agree not to publish or use in presentations UCF specific network information; e.g., IP addresses, labels, URLs, etc.

Data users must have a bona fide academic or UCF administrative reason to request network data

Project Procedures and Outline:

Data user(s): ____*Susan E. McGill*____________

Duration of Project: ____*Dec. 2005 – Aug. 2007*__

Proposed Use - This should be a summary of the research project or the administrative purpose for which the data will be used:

Modeling and Simulation Doctoral Dissertation -This project will result in a simulation modeling portions of the UCF campus network. So many details will be contained that at the end of the project the simulation must be turned over to IT for archiving. It will be released for future research as authorized.

Collect Configuration and Traffic data for an OPNET simulation.

Data Collection Team: ____
Name(s)

Format of the Data: ____

Location of Stored Data: ____

Acknowledgment

By signing below, you agree to the terms outlined above:

Data User: ____
Sign Date

Data Collection Team: ____
Sign Date

IT Security: ____
Sign Date

Director: ____
Sign Date

Glossary:

Data users: Researcher requesting access to data collected by the data collection team.

Data Collection Team: Staff of Computer Services charged with the authority to collect network data streams for the purpose of research or investigation.

NOC: Network Operations Center – Division of Computer Services and Telecommunication.

UCF Reference Standards: Data Classification Standard, IT Security Policy, Administrative Data, Information, and Computer Security Policy, Administrative Data, Information, and Computer Security Guidelines.

***The actual authorization form and image of the hard drive used for this research are on file with UCF Information Resources IT Security**

LIST OF REFERENCES

[1] S. Aidarous and T. Plevyak, *Managing IP networks: challenges and opportunities*. Hoboken, NJ: John Wiley, 2003.

[2] D. Messerschmitt, *Making IT better: Expanding Information Technology Research to Meet Society's Needs*. Washington, D.C: National Academy Press, 2000.

[3] U. Nulden, "Escalation in IT projects: can we afford to quit or do we have to continue?" 1996.

[4] I. R. Furmston-Evans, "Successful system architecture development for enterprise wide client server implementation," 1995.

[5] D. Patterson and Committee, *Looking Over the Fence at Networks: A Neighbors View of Networking Research*. Washington, D.C.: National Academy Press, 2001.

[6] W. B. Rouse and M. L. Baba, "Enterprise transformation," *Communications of the ACM*, vol. 49, pp. 66-72, 2006.

[7] W. E. Bijker, *Of Bicycles, Bakelites, and Bulbs: Toward a Theory of Sociotechnical Change*. Cambridge, Ma.: The MIT Press, 1995.

[8] R. Kling and R. Lamb, "IT and organizational change in digital economies: a socio-technical approach," *SIGCAS Comput. Soc.*, vol. 29, pp. 17-25, 1999.

[9] M. L. Tushman and P. Anderson, *Manageing Stategic Innovation and Change: Selected Readings*. New York: Oxford University Press, 1997.

[10] A. Oodan, K. Ward, C. Savolaine, M. Daneshmand, and P. Hoath, *Telecommunications Quality of Service Management: from legacy to emerging services*. London, UK: Institute of Electrical Engineers, 2003.

[11] D. L. a. H. Radha, "Retransmission schemes for streaming internet multimedia: evaluation model and performance analysis," *SIGCOMM Comput. Commun. Rev*, vol. 32, pp. 70--83, 2002.

[12] Z. Ming, B. Karp, S. Floyd, and L. Peterson, "RR-TCP: a reordering-robust TCP with DSACK," 2003.

[13] M. Siller and J. C. Woods, "QoS arbitration for improving the QoE in multimedia transmission," presented at International Conference on Visual Information Engineering, 2003.

[14] J. Laiho, K. Raivio, P. Lehtimaki, K. Hatonen, and I. Simula, "Advanced analysis methods for 3G cellular networks," *Ieee Transactions On Wireless Communications*, vol. 4, pp. 930-942, 2005.

[15] A. Dodd, *The Essential Guide to Telecommunications*, 3 ed. Upper Saddle River, N.J.: Prentice-Hall, 2002.

[16] V. Hlupic, "Simulation software: an Operational Research Society survey of academic and industrial users," presented at Simulation Conference Proceedings, 2000. Winter, 2000.

[17] D. G. Messerschmitt, *Understanding Networked Applications: A First Course*. San Francisco, Ca.: Morgan Kaufmann Publishers, 2000.

[18] S. Shepard, *Telecommunications Convergence: How to bridge the gap between technologies and services*, 2 ed. New York: McGraw-Hill, 2002.

[19] H. J. Fowler, "TMN-based broadband ATM network management," *Communications Magazine, IEEE*, vol. 33, pp. 74, 1995.

[20] CCITT and ITU, "Recommendation X711: Common Management Information Service Definition for CCITT Applications," 1991.

[21] W. Stallings and W. Stallings, *SNMP, SNMPv2, SNMPv3, and RMON 1 and 2*, 3rd ed. Reading, Mass.: Addison-Wesley, 1999.

[22] A. Cerra and D. Kan, "User-centric broadband services: Demand drivers and market opportunities," *Alcatel Telecommunications Review*, pp. 6-10, 2005.

[23] I. Motorola, "Motorola Networks - New and Expanded Business," 2005.

[24] W. Willinger, R. Govindan, S. Jamin, V. Paxson, and S. Shenker, "Scaling phenomena in the Internet: Critically examining criticality," *Proceedings Of The National Academy Of Sciences Of The United States Of America*, vol. 99, pp. 2573-2580, 2002.

[25] J. Fitzgerald and A. Dennis, *Business Communications and networking*, 7 ed. New York: John Wiley & Sons, Inc, 2002.

[26] K. Claffey, "Internet Traffic Characterization," in *Computer Science and Engineering*. San Diego: University of California, San Diego, 1994, pp. 121.

[27] J. S. Strassner, *Directory enabled networks*. Indianapolis, IN: Macmillan Technical, 1999.

[28] J. van Bogaert, Y. T'Joens, and J. P. Lartigue, "Fixed access vision," *Alcatel Telecommunications Review*, pp. 140-143, 2005.

[29] J. Kumagai, "Mission impossible? [FBI computer network]," *Spectrum, IEEE*, vol. 40, pp. 26, 2003.

[30] L. Dong-Ho, L. Sang-Rok, Y. Byung-Ho, C. Moo-Jung, and K. Jae-Keun, "In-service performance monitoring of optical amplifiers utilizing band-rejection filter," 1999.

[31] G. Kotsis, "Performance management in dynamic computing environments," in *Performance Tools And Applications To Networked Systems*, vol. 2965, *Lecture Notes In Computer Science*. Berlin: Springer-Verlag Berlin, 2004, pp. 254-264.

[32] B. L. Tierney, W. E. Johnston, J. R. Lee, and G. Hoo, "Performance analysis in high-speed wide area IP-over-ATM network: top-to-bottom end-to-end monitoring," *Network, IEEE*, vol. 10, pp. 26, 1996.

[33] L. L. Qiu, Y. Zhang, and S. Keshav, "Understanding the performance of many TCP flows," *Computer Networks-The International Journal Of Computer And Telecommunications Networking*, vol. 37, pp. 277-306, 2001.

[34] IEC, "Performance Management for Next-Generation Networks," 2005.

[35] CiscoSystems, "Cisco Catalyst 6500 Series Network Analysis Module (NAM-1/NAM-2)," vol. 2006, 2006.

[36] F. Krief, "Self-aware management of IP networks with QOS guarantees," *Int. J. Netw. Manag.*, vol. 14, pp. 351-364, 2004.

[37] M. Parashar and S. Hariri, "Autonomic computing: An overview," in *Unconventional Programming Paradigms*, vol. 3566, *Lecture Notes In Computer Science*. Berlin: Springer-Verlag Berlin, 2005, pp. 257-269.

[38] T. M. Chen and S. S. Liu, "A model and evaluation of distributed network management approaches," *Selected Areas in Communications, IEEE Journal on*, vol. 20, pp. 850, 2002.

[39] J. Strassner, *Policy-Based Network Management: Solutions for the Next Generation*. San Francisco: Morgan Kaufmann Publishers, 2004.

[40] J. Strassner, "Autonomic networking - theory and practice," 2004.

[41] A. M. Law and W. D. Kelton, *Simulation Modeling and Analysis*, 2nd ed. New York: McGraw-Hill, Inc., 1991.

[42] J. W. Schmidt and R. E. Taylor, *Simulation and Analysis of Industrial Systems*. Homewood, Il: Richard D. Irwin, 1970.

[43] A. Medina, M. Allman, and S. Floyd, "Measuring the evolution of transport protocols in the Internet," *Computer Communication Review*, vol. 35, pp. 37-51, 2005.

[44] B. Sikdar, S. Kalyanaraman, and K. S. Vastola, "Analytic models for the latency and steady-state throughput of TCP Tahoe, Reno, and SACK," *Ieee-Acm Transactions On Networking*, vol. 11, pp. 959-971, 2003.

[45] G. Kaplan, "Simulating networks," *Spectrum, IEEE*, vol. 38, pp. 74, 2001.

[46] W. D. Kelton, R. P. Sadowski, and D. T. Sturrock, *Simulation with Arena*, 3rd ed. New York: McGraw Hill, 2004.

[47] S. Floyd and V. Paxson, "Difficulties in simulating the Internet," *Networking, IEEE/ACM Transactions on*, vol. 9, pp. 392, 2001.

[48] R. R. Iraschko and W. D. Grover, "A highly efficient path-restoration protocol for management of optical network transport integrity," *Ieee Journal On Selected Areas In Communications*, vol. 18, pp. 779-794, 2000.

[49] S. Floyd, M. Handley, J. Padhye, and J. Widmer, "Equation-based congestion control for unicast applications," *Computer Communication Review*, vol. 30, pp. 43-56, 2000.

[50] S. I. Maniatis, E. G. Nikolouzou, and I. S. Venieris, "End-to-end QoS specification issues in the converged MAP wired and wireless environment," *Ieee Communications Magazine*, vol. 42, pp. 80-86, 2004.

[51] G. Schorcht, I. Troxel, K. Farhangian, P. Unger, D. Zinn, C. K. Mick, A. George, and H. Salzwedel, "System-level simulation modeling with MLDesigner," 2003.

[52] D. M. Moen, J. M. Pullen, and Z. Fei, "Implementation of Host-Based Overlay Multicast to Support of Web Based Services for RT-DVS," 2004.

[53] J. Dowling, E. Curran, R. Cunningham, and V. Cahill, "Using feedback in collaborative reinforcement learning to adaptively optimize MANET routing," *Ieee Transactions On Systems Man And Cybernetics Part A-Systems And Humans*, vol. 35, pp. 360-372, 2005.

[54] K. Jain, J. Padhye, V. N. Padmanabhan, and L. L. Qiu, "Impact of interference on multi-hop wireless network performance," *Wireless Networks*, vol. 11, pp. 471-487, 2005.

[55] A. E. Eshoul and H. T. Mouftah, "A computer aided tool for the performance evaluation of next generation WDM wavelength-routed optical networks," *Computer Communications*, vol. 27, pp. 1197-1205, 2004.

[56] T. B. Kuang and C. Williamson, "Hierarchical analysis of RealMedia streaming traffic on an IEEE 802.11b wireless LAN," *Computer Communications*, vol. 27, pp. 538-548, 2004.

[57] I. Motorola, "NetPlan," 2002, pp. http://www.motorola.com/mot/doc/1/1336_MotDoc.pdf.

[58] K. Pawlikowski, H. D. J. Jeong, and J. S. R. Lee, "On credibility of simulation studies of telecommunication networks," *Communications Magazine, IEEE*, vol. 40, pp. 132, 2002.

[59] S. Kurkowski, T. Camp, and M. Colagrosso, "MANET simulation studies: the incredibles," *SIGMOBILE Mobile Computing Communication Review*, vol. 9, pp. 50-61, 2005.

[60] T. R. Andel and A. Yasinsac, "On the credibility of manet simulations," *Computer*, vol. 39, pp. 48, 2006.

[61] G. F. Lucio, M. Paredes-Farrera, E. Jammeh, M. Fleury, and M. J. Reed, "OPNET Modeler and Ns-2: Comparing the Accuracy Of Network Simulators for Packet-Level Analysis using a Network Testbed," vol. 2004: Proceedings 3rd WEAS Int. Conf. on Simulation, Modeling and Optimization (ICOSMO 2003), vol. 2, pp. 700–707, Crete, October 2003., 2004.

[62] C. Marchetti, B. Pernici, and P. Plebani, "A Quality Model for Multichannel Adaptive Information Systems," presented at WWW2004, New York, New York, USA, 2004.

[63] R. Beuran, M. Ivanovici, B. Dobinson, N. Davies, and P. Thompson, "Network Quality of Service Measurement System for Application Requirements Evaluation," presented at SPECTS'03, 2003.

[64] G. J. Armitage, "Revisiting IP QoS: Why do we care, What have we learned? ACM SIGCOMM 2003 RIPQOS Workshop Report," *ACM SIGCOMM Computer Communications Review*, vol. 33, pp. 81-88, 2003.

[65] P. Quax, Monsieurs, P., Lamotte, W., De Vleeschauwer, D., and Degrande, N., "Objective and subjective evaluation of the influence of small amounts of delay and jitter on a recent first person shooter game.," presented at 3rd ACM SIGCOMM Workshop on Network and System Support For Games, Portland, Oregon, USA, 2004.

[66] J. Nichols and M. Claypool, "The effects of latency on online madden NFL football," presented at NOSSDAV '04 workshop Network and Operating Systems Support For Digital Audio and Video, Cork, Ireland, 2004.

[67] J. But, "The brave new world of online digital home entertainment," *IEEE communications magazine*, vol. 43, pp. 84, 2005.

[68] P. Calyam, M. Sridharan, W. Mandrawa, and P. Schopis, " Performance Measurement and Analysis of H.323 Traffic," presented at PAM'04, 2004.

[69] S. R. Gulliver and G. Ghinea, "Defining User Perseption of Distributed Multimedia Quality," *ACM Transactions of Multimedia Computing, Communications and Applications*, vol. 2, pp. 241-257, November 2006.

[70] D. E. McDysan, *QoS & traffic management in IP & ATM networks*. New York: McGraw Hill, 2000.

[71] E. Yaprak, "Developing a Framework for Effective Network Capacity Planning," NASA/JSC, Communications Branch, Information Technology Division July 9 2004 2004.

[72] K. Pawlikowski, H. D. J. Jeong, and J. S. R. Lee, "On credibility of simulation studies of telecommunications networks," *Communications Magazine, IEEE*, vol. 40, pp. 132, 2002.

[73] OPNET, *OPNET Modeler User Guide*, Release 11.0 ed: Available only through OPNET, 2005.

[74] T. Szigeti and C. Hattingh, *End-to-end qos network design*. Indianapolis, IN: Cisco Press, 2005.