Interaction between Multimedia and Traditional (FTP/Web) Applications: An Experimental Performance Study

Gurvinder Singh and John Silvester

CENG 98-24

Department of Electrical Engineering - Systems
University of Southern California
Los Angeles, California 90089-2562
(213-740-4579)
December 1998

Interaction between Multimedia and Traditional (FTP/Web) Applications: An Experimental Performance Study

Gurvinder Singh and John Silvester
University of Southern California
Department of Electrical Engineering-System
Los Angeles, CA 90089-2562

John Vicente, FM1-152
Intel Corporation
1900 Prairie City Road
Folsom, CA 95630

October 1997

Abstract

In this paper, we present the results from network characterization experiments performed with Intel Corporation multimedia applications. We explain these results as practical evidence of the behavioral impacts of multimedia applications against traditional applications, specifically FTP file and Web-based transfer activity which behave under the auspices of TCP/IP. Utilizing statistical performance analysis techniques, we characterize a set of multimedia applications and analyze them for network scaling under isolated and mixed (with FTP/Web activity) traffic conditions. We call out specific issues related to this study, prevalent in today's network environments that have contributed to the delayed introduction of these modern applications. We present a realizable characterization and network planning methodology to address this complex issue. Finally, we provide suggestions for further work or analysis that may aid multimedia planners in the introduction and transition of multimedia services to today's service-level driven and "reduce cost of computing" environment.

1. Introduction

The basic Internet offers two classes of transport protocols UDP and TCP. The User datagram protocol (UDP) is an unreliable connectionless protocol. It has no sequence number and error recovery mechanism. It is a best effort service provided by IP, and is used mainly for applications which do not require reliable data transfer e.g., voice and video. On the other hand, Transmission Control Protocol (TCP) is a reliable connection-oriented protocol. It is stream-oriented and supports duplex message transfer services. TCP incorporates Go Back N (GBN) error control using adaptive sliding window flow control method. Buffer size, distance, network load and link capacities are some of the key factors that influence the window size.

In today's corporate environment time-critical (traditional) applications and service levels are important to running a successful business. Introducing new UDP-based voice/video applications or integrated multimedia applications can have a major impact in maintaining these service levels. The two types of traffic on which we are focusing, *multimedia* traffic (integrated voice, video & data; mainly UDP) and *traditional* traffic (FTP, Telnet, SMTP, HTTP; mainly TCP) can both coexist with stability and minor interaction provided sufficient bandwidth is available. Nevertheless, as we increase UDP traffic on a given segment or network link, TCP or the traditional traffic will back off and tend to suffer. We observe both the isolated (multimedia traffic only) and mixed traffic implications as part of the multimedia network capacity planning process -- to analyze how many multimedia application streams can coexist without affecting the traditional applications' performance. Over-provisioning of the network is one solution; it does provide a safe, but costly alternative. Moreover, the term "severe increase in delay" is subjective. For example, consider an FTP file transfer -- an increase in delay of 100% can be viewed as a major service impact if a normal 20-minute FTP transfer increases to 40 minutes due to the introduction of these multimedia applications. On the other hand, a 1 second to 2 second increase when the same traffic is introduced may fall well within the necessary service levels. This paper analyzes some of these issues directly as we present the measurement analysis and methodology and some of our key findings.

The second section describes the applications and tools used in the measurement experiments. We provide a brief description of the protocols and encoding specifications. In section (3), the selected network topologies are described. We discuss the test cases and provide some background in choosing these test scenarios including our trace analysis, performance evaluation and detailed analysis of the multimedia applications in section (4). In section (5) and (6), we present the impact and scalability analysis, respectively, of the multimedia applications and their interaction with the traditional applications. Finally, in the last section, we summarize our findings, preliminary conclusions, and suggestions for further work.

2. Applications Studied

- **PRODUCT Z** is a Web-base distance learning/training, application. It has the ability to multicast a presentation live to an audience, and/or store it on disk so it may be played back (unicast, on-demand). This application consists of a camera and associated capture hardware and software. The display contains an AVI control, window, an area for displaying what the camera captures, and a presentation area. Any content that can be displayed by a Web browser can be displayed in the presentation area. The application runs embedded in an ActiveX enabled Web browser. It is ActiveMovie based and was designed to use ActiveRTP as the transport for streaming multimedia. It uses G.723 for audio and H. 263 for the Video Codec. **ToolMM**, a video streaming application was used in our test cases to simulate the PRODUCT Z application.

- **PRODUCT X** is a point-to-point (unicast) video conferencing tool. It can be used on LANs, WANs and ISDN. It uses IP/UDP complemented by the RSVP/RTP protocols, which provide guaranteed quality of service and bandwidth when implemented in an IP environment. It also uses H.225.0 signaling, G.723 for audio and H.263 for the Video Codec.

- **PRODUCT Y** is Precept's desktop video broadcasting/multicasting solution. It uses RTP/RTCP, H.261 video and various audio codecs. It is compatible with Unix VIC / VAT. It uses existing data networks and offers a sophisticated client/server application for enterprise video broadcasting/multicasting. Network bandwidth is conserved though the use of IP multicasting, PRODUCT Y also allows the choice of various compression and decompression (codec) schemes to ensure optimal picture quality depending on the specific application and network bandwidth availability.

- **FTP** (File Transfer Protocol) In this application, the file server can be accessed by single or multiple clients, as the server interacts with the local file system to carry out the requested operation. It supports three file types -- unstructured, structured and random access. Typically, two TCP connections are established for each session. One for exchanging the control messages and the other for transferring files. In this transaction scenario, the file delivery must be complete before the user can do anything with it. The overall file transfer time is the performance focus.

- The **World Wide Web (WWW)** is used to access objects using an extendable number of protocols. **HTTP** has been in use by the WWW global information initiative since 1990. It is an application-level, generic, stateless, and object-oriented protocol for distributed, collaborative, hypermedia information systems. A client sends a request to the server in the form of a request method, URL, and protocol version. It is followed by a MIME (Multipurpose Internet Mail Extensions) -like message containing request modifiers, client information, and possible body content over a connection with a server. The server responds with a status line, including the message's protocol version and a success or error code, followed by a MIME-like message containing server information, entity meta-information and possible entity-body content. HTTP communication usually runs over TCP/IP connections. In HTTP/1.0, most implementations used a new connection for each request/response exchange.

- **ToolGEN** is a 'home-grown' Unicast /Multicast Network Performance measurement tool set. Script files can be written to generate real time traffic patterns so that the network can be loaded in a variety of ways and the loading pattern can be made to change over the course of time. One can emulate the traffic patterns of unicast or multicast UDP/IP applications. It can be used to dynamically join and leave IP multicast groups and log the received data. It can also be used to calculate the statistics on received data such as throughput, packet loss and communication delay.

3. Network Topologies

Figure 1: A simple Ethernet topology for capturing and analyzing multimedia traffic streams.

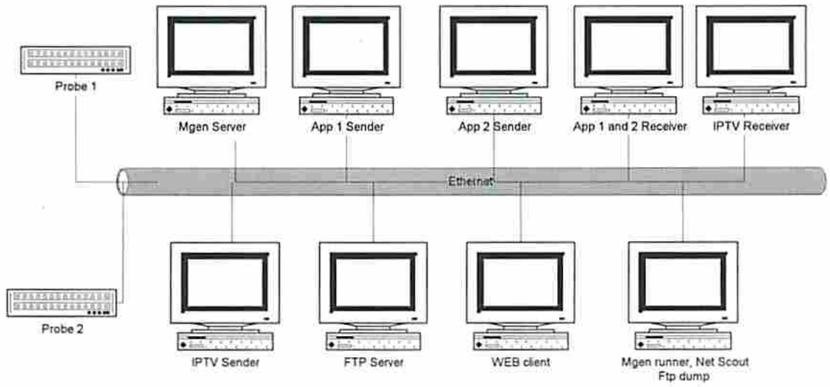


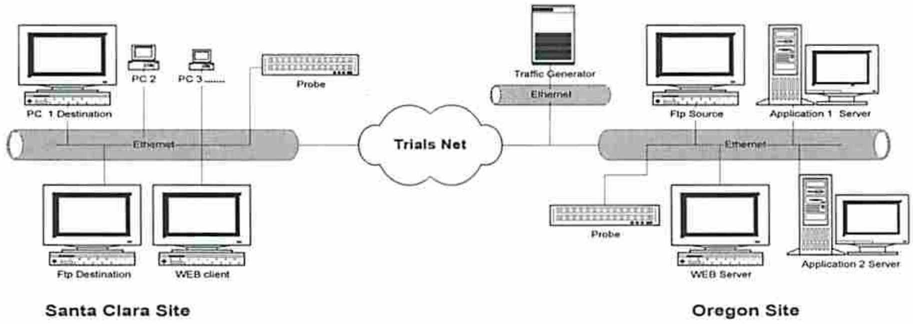
Figure [1] shows a controlled, 'pure' 10Mbs Ethernet topology with no external variables; the controlled bottleneck being the bandwidth of the Ethernet. The 'pure' case was our baseline configuration to comparatively scale with more complex topologies. All clients and servers were on the same LAN except for the Web servers. The path to the Web servers was validated to not be a delay bottleneck in these experiments. Moreover, we did not use proxy servers or allow desktop caching, causing the web client to directly access the specific Web servers (script-driven) iteratively rather than loading it from cache.

Our primary objectives for this configuration were to capture traces for different applications at predefined utilization levels and analyze their behavior and impacts. To do this, we generated CDF graphs, bandwidth utilization and the autocorrelation functions for all test cases. Through trace analysis, we could classify the applications as 'hard' or rate adaptive.

Additionally, a live WAN (Figure [2]) test scenario was chosen; here our objective was to test the application on a live production WAN and compare the performance results with that of the simple Ethernet, our baseline case. In this environment, we tested both multicast/unicast applications. With clients and servers dispersed on both sides (Santa Clara, Oregon) of the WAN Trials network.

Figure 2:

The real WAN test scenario for both unicast and multicast Environment



For appropriate scope and brevity, we have concentrated our focus primarily on the test results associated with Figure [1] - the Ethernet case in this paper. Our key points and findings are based on the Ethernet analyses as we defer our comparative WAN results for a future publication. In summary, we focus our attention to (a) behavioral, interactive impacts between multimedia and traditional applications, and (b) towards developing a performance methodology of the multimedia applications for further network scaling and comparative analysis.

4. Measurement Experiments and Trace Analysis

We report on the 16 test cases with different utilization levels on the Ethernet LAN shown in Table [1]. The parameters used for different multimedia applications are presented in Table [2] and sample utilization graphs are shown in Figures [3,4,5]. A traffic generator, ToolGEN, was used to vary overall Ethernet utilization. For example, test cases 2 and 3 are for the same application mix but with additional background of 4.2Mbps from ToolGEN. The ToolGEN scripts were written with an objective (and assumption) to generate ToolGEN traffic equivalent to multiple multimedia streams.

Table 1: LAN Test Cases: Application Mixes (# Sessions of each Application Type).

	Product Z -1	Product Z –2	Product Z - 3	Product Y	Product X	Web (HTTP)	FTP	ToolGEN	Collision %
Test 1	1	1	0	0	0	0	0	0	0.1 %
Test 2	1	1	0	0	0	1	4	0	1.5 %
Test 3	1	1	0	0	0	1	4	4.2Mbps	18 %
Test 4	1	1	1	0	0	1	4	6.8Mbps	25 %
Test 5	1	1	1	0	0	1	4	7.6Mbps	25 %
Test 6	1	0	1	0	0	1	4	0	0.5 %
Test 7	1	0	1	0	0	1	4	3.6Mbps	5 %
Test 8	0	0	0	2	0	0	0	0	0.2 %
Test 9	1	0	0	2	0	0	0	0	0.2%
Test 10	1	0	0	2	1	0	0	0	0.5 %
Test 11	0	0	0	2	1	1	4	0	10 %
Test 12	0	0	0	2	1	1	4	3.6Mbps	17 %
Test 13	0	0	0	2	1	1	2	6.0Mbps	20 %
Test 14	0	0	0	2	1	1	2	2.4Mbps	14 %
Test 15	0	0	0	2	1	1	2	1.6Mbps	10 %
Test 16	0	0	0	2	1	1	2	4.4Mbps	16%

Table 2: Application Parameters

Application	TYPE	Parameters Used
Product Z - 1	AVI file	CIF 8 fps
Product Z - 2	LIVE	QCIF (128 ,8543,4,12 fps)
Product Z - 3	LIVE	CIF (128 ,8543,4,12 fps)
Product X	LIVE	Highest Quality
Product Y	LIVE	H.261, 24 bit (320, 240)

➢ *Frames Per Sec (fps)*

Figure 3: PRODUCT Y Bandwidth Utilization Graph

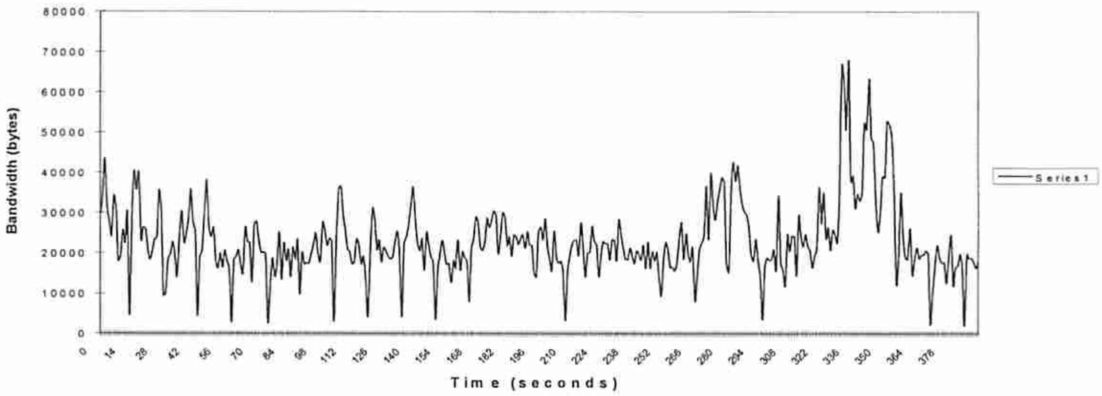


Figure 4: PRODUCT X Bandwidth Utilization Graph

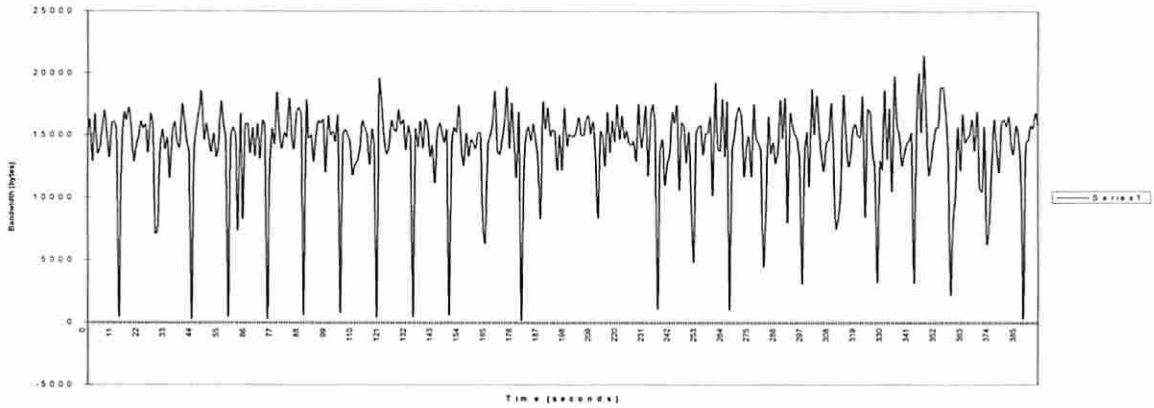
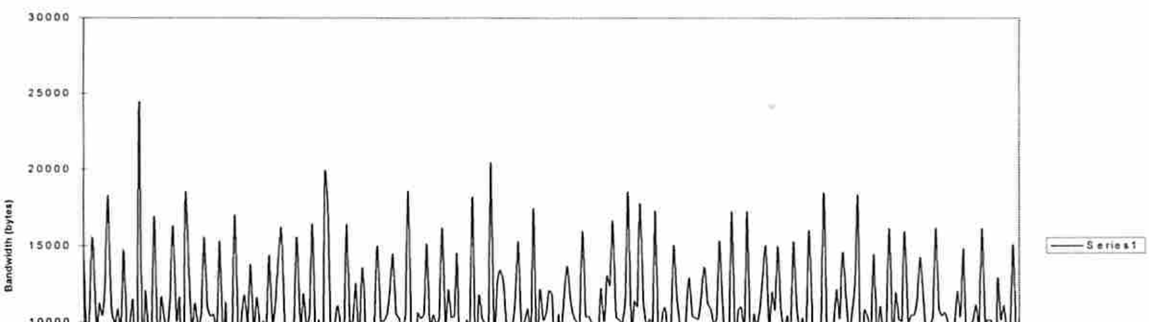


Figure 5: PRODUCT Z Bandwidth Utilization Graph



As mentioned earlier, ToolMM corresponds to the video portion of the PRODUCT Z application, and most of the test cases were run 'live' with QCIF and CIF quality for ToolMM. With increased utilization levels and collisions, the TCP-based traffic backed-off and TCP applications get reduced throughput. On the other hand, there was no effect on the multimedia application's video stream quality even when we increased the utilization to 80%! At very high utilization levels, the 'live' capture quality was 'poorer' compared to the recorded AVI file transmission. Our conclusion was that the AVI file had already been compressed and streamlined as compared to the 'live' transmission stream. In the 'live' case, the process had to grab the frame, encode, apply compression and finally streamline the video prior to transmission.

Our experiments were performed to observe the changes in the applications' traffic patterns under no load, medium load and heavy load conditions. Tests were also conducted with the multimedia applications in isolation without any other applications on the network. In general, any change in the overall traffic scene usually caused a minor change in traffic profile by the live multimedia applications. PRODUCT X network utilization was fairly constant (see below) with slight utilization variations due to the "whiteboard' activity, a product feature. Both HTTP and FTP were directly affected by the ToolGEN background traffic used to simulate an increasing, controlled collision environment. With each established utilization level and the corresponding adjustments in the TCP window sizes, throughput fluctuations were dramatic for FTP/Web, but eventually stabilizing due to TCP stability. We discuss this impact in more detail in the next section on impact analysis.

The autocorrelation function is calculated and illustrated in Figures [6,7,8], for each of the multimedia applications. For PRODUCT X, the arrivals are positively correlated for all lags less than 500. This correlation indicates that PRODUCT X may require more bandwidth than its average for moderately long periods of time. In the case of PRODUCT Y, the autocorrelation is negative; suggesting the arrivals to be negatively correlated for all lags less than 500. For negatively correlated arrivals, the impact on the network performance evaluation is not clear; further analysis is required on this issue. The PRODUCT Z graph shows zero autocorrelation indicating the traffic pattern is independent in nature. That is, the traffic is well behaved and an allocation of average rate may be sufficient for supporting this application with a desired QoS.

Figure 6: PRODUCT X - 10 msec samples

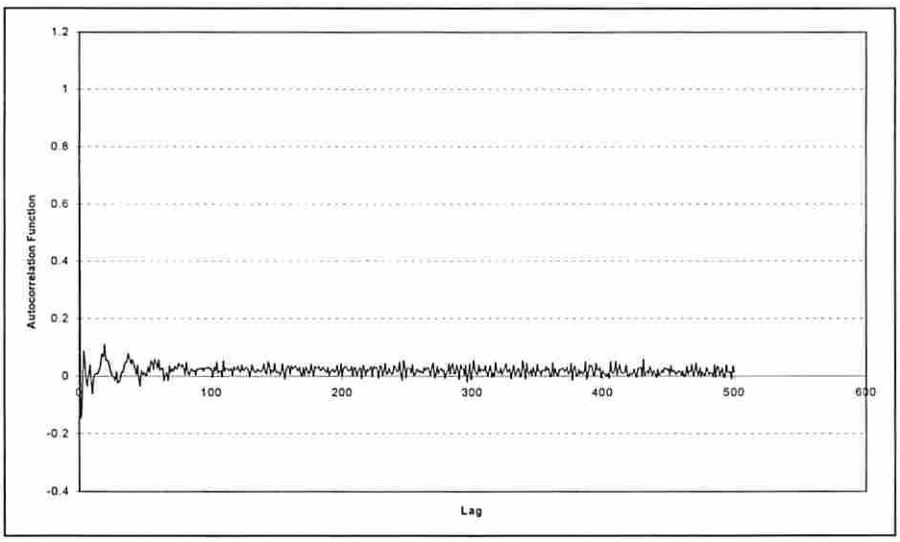


Figure 7: PRODUCT Y - 10 msec

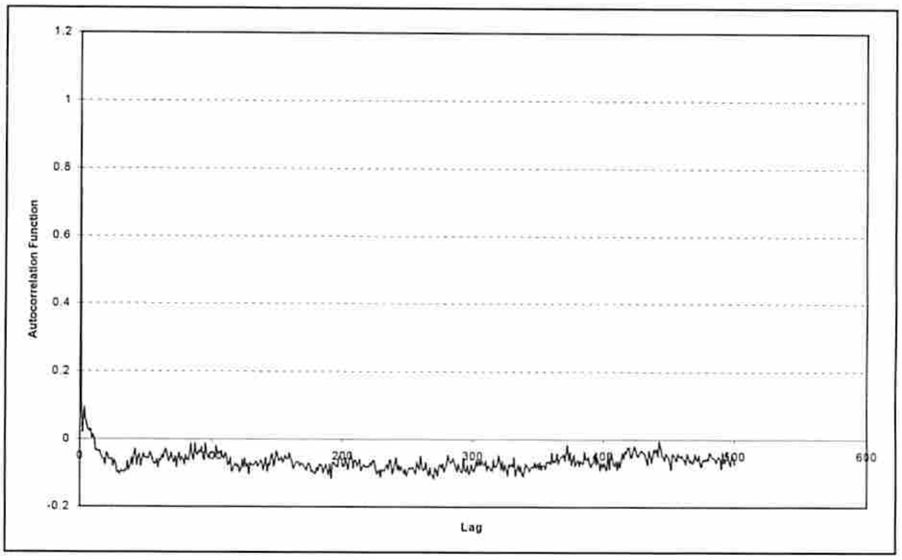
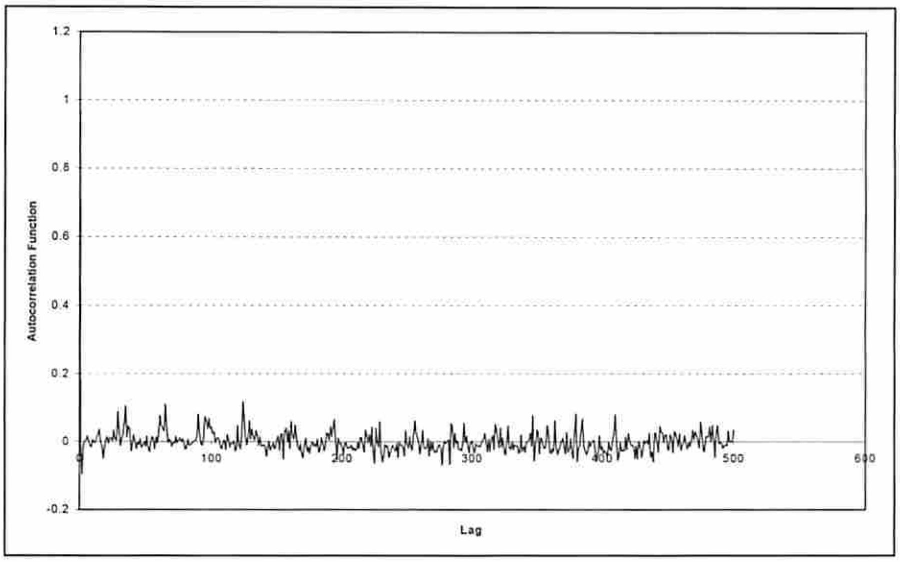


Figure 8: PRODUCT Z - 10 msec



In Figures [9,10,11] the CDF (Cumulative Distribution Function) graphs depict the application throughput v. probability. We found that the CDF curves for the applications did not vary much from one test case to another. Average values of the tests are shown in Table [3]. The dotted line crosses the y-axis at 0.7 probability and corresponding x-axis on the measured throughput of the application. This says that 70% of the time, the application was using less than the associated (x-axis) throughput value. Our recommended throughput assignment is 70%; the assumption being that the multiplexing gain will compensate for the peaks in the utilization occurring above this value. Thus, we are suggesting neither allocating peak rate (leading to over-provisioning) to the given application and nor the average rate (under-provisioning). This value then becomes the methodology basis for scaling calculations for individual multimedia application streams. The 70% CDF value for PRODUCT Y on the average was found to be 212Kbps. The fluctuation in the 70% CDF value over the range of tests performed was not more than 5 to 6%. Therefore, we can take this value and say that the bandwidth consumed by PRODUCT Y was less than 212 Kbps, 70 % of the time. Similarly the 70% CDF bandwidth value for PRODUCT X was found out to be 128.8Kbps. In the PRODUCT Z case, for various parameters [PRODUCT Z1, PRODUCT Z2, PRODUCT Z 3] we had different CDF values. In all cases, the tests show these applications were not affected by the increase in utilization of the Ethernet, indicating 'hard' and not rate-adaptive behavior.

Table 3: Cumulative Distribution Function (CDF)

APPLICATIONS	Product Z 1	Product Z 2	Product Z 3	Product X	Product Y
0%	12816	6080	52800	30400	21120
50%	75784	78560	304000	120800	168800
70%	92656	88800	336000	128800	212000
90%	137008	104720	396000	152000	274400
100%	232000	192000	469728	231200	549360

➢ *Note values are in bits/sec*

Figure 9: Product Z CDF

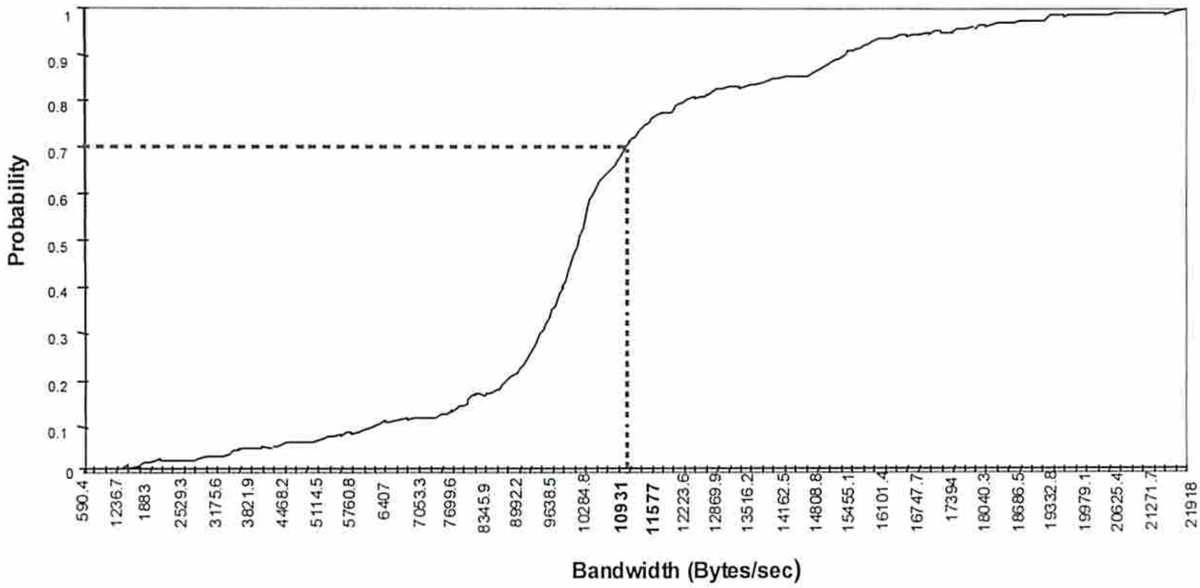


Figure 10: Product X CDF

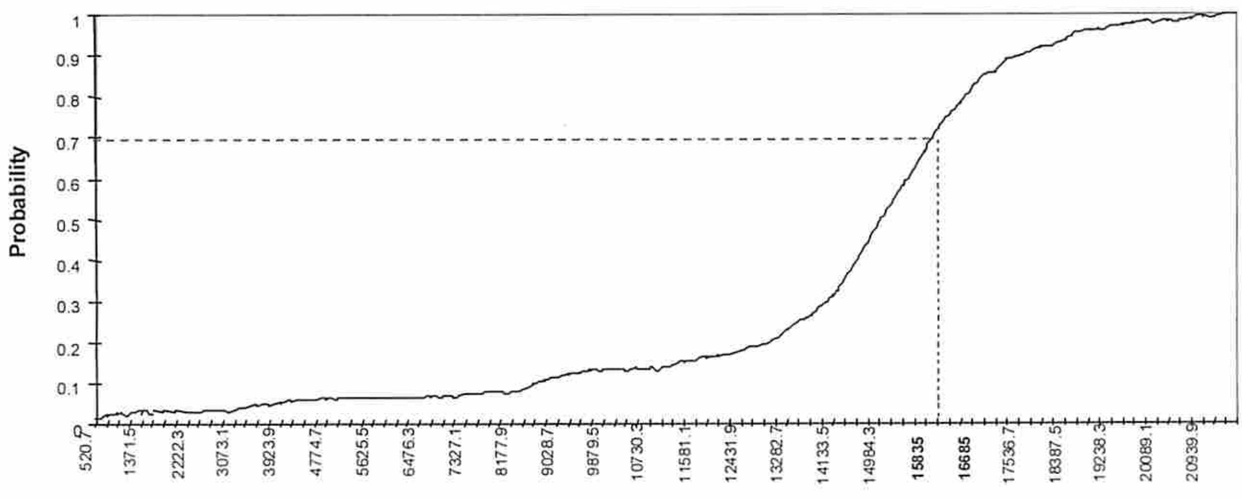
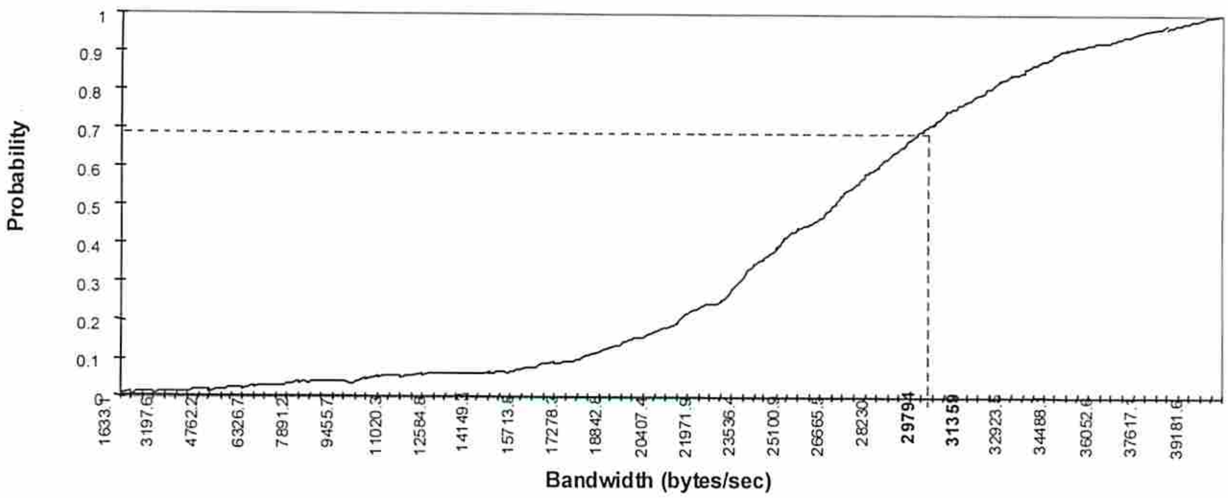


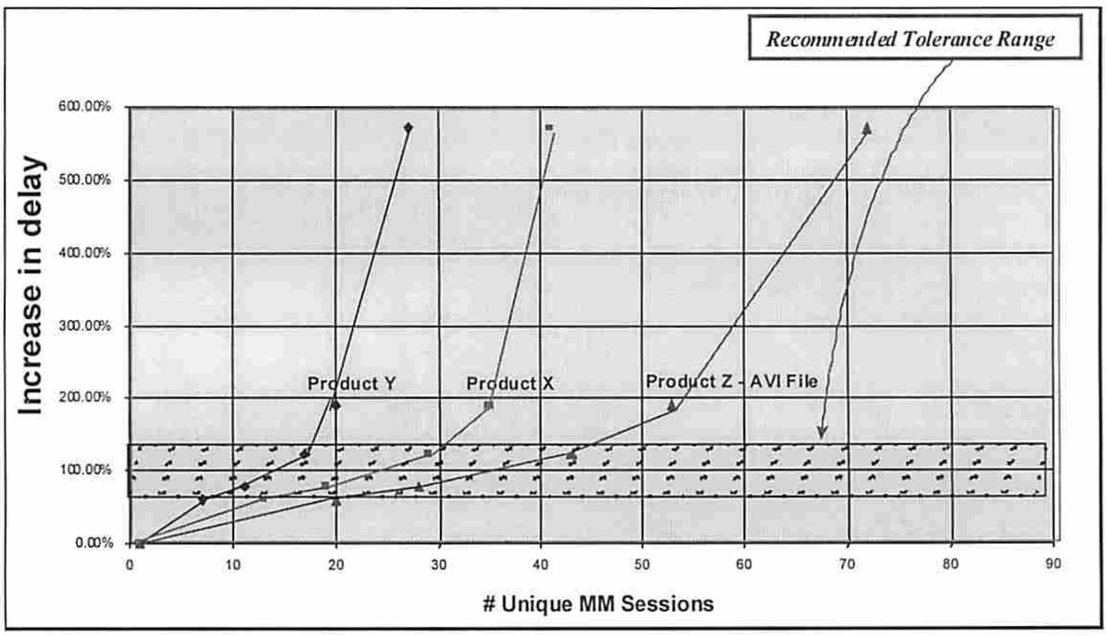
Figure 11: Product Y CDF



5. Impact Analysis on Traditional Applications (FTP, HTTP, Telnet)

The impact analysis graphs Figures [`12] and [13] depict the increase in delay experienced by the FTP and WWW, respectively, as we increase the number of unique multimedia sessions for different applications. We see from the graph that the impact on FTP is nonlinear. The points on these graphs correspond to different test cases for FTP (HTTP) and the overall utilization variation induced by ToolGEN. The number of unique (x-axis) sessions of an application is calculated by converting the ToolGEN load into an equivalent number of distinct application sessions based on 70% CDF values. The number of distinct sessions could imply individual multicast groups initiated or unicast sessions and is NOT the number of users. The multicast nature makes these applications more scalable than the unicast application. In either case, as we increase the number of distinct sessions of the given multimedia application, the increase in delay is nonlinear. The graphs show that there is a knee point for each application. The behavior is similar for the HTTP (WWW) application, but the increase in delay is not as dramatic as for FTP. As a reminder, proxy service and local caching were disabled during our testing to eliminate these potential bottleneck variables. The reason for the lower increase in HTTP delay can be attributed to the page size which resulted in a transfer which was generally smaller than the FTP transfer that does not allow for much adaptivity from TCP.

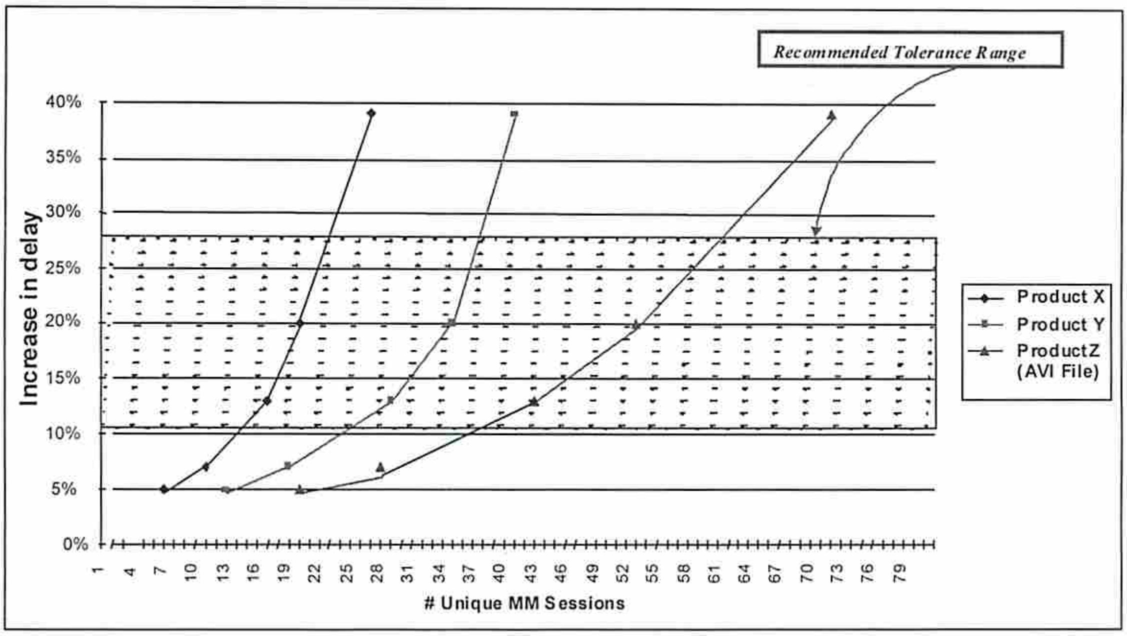
Figure 12: Graphs of FTP



** Note the Number of sessions for PRODUCT Y and for PRODUCT Z are just for the video streams.*

Most of our calculations and analysis are dependent on the assumption that the traffic generated by the ToolGEN simulator is a reasonable approximation to the collective application behavior. The script-based traffic simulator (ToolGEN) generated a continuous, collision simulated traffic profile. Although the ToolGEN simulator did not generate the identical inter-arrival traffic as the original multimedia applications, the behavior was continuous and non-adaptive, with similar average rate and peak periods. We contend that the impact of this former assumption (inter-arrival discrepancy) will not make a significant variance in the test results, as the other traffic attributes are key to driving these same results. The tolerance band shown is a possible range for acceptable performance degradation for the traffic – but this clearly depends on how the application is used.

Figure 13: Graphs of HTTP

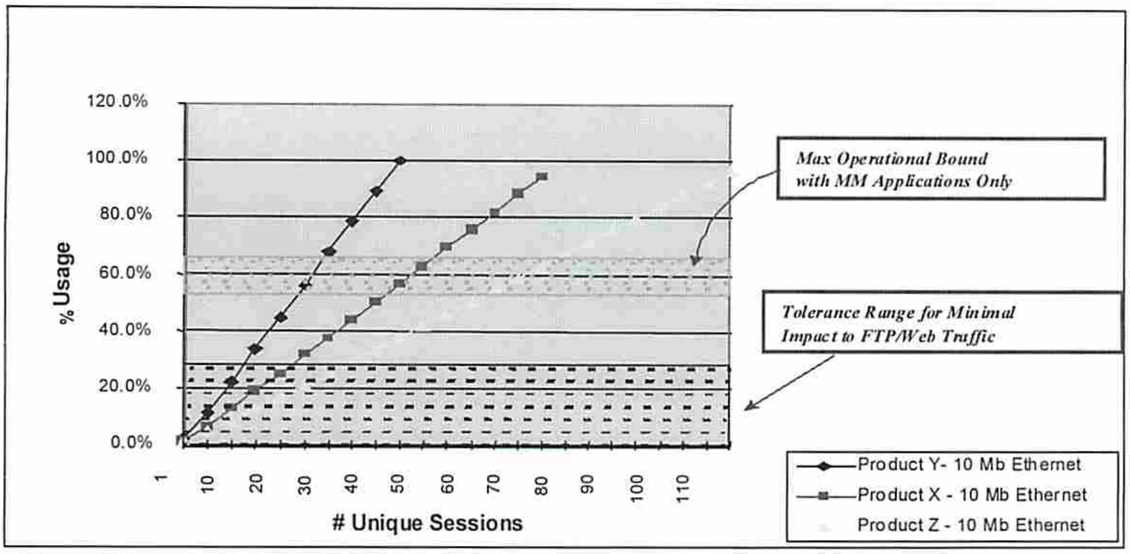


6. Multimedia Application Scaling Analysis

In Figure [14], the projected values are calculated with respect to the 70% CDF values. The three rising lines shown in Figure [14] are for PRODUCT Y, PRODUCT X and PRODUCT Z. Recommended operating and scaling ranges are also shown. We recommend two ranges: the lower range is where both TCP traffic and the Multimedia traffic must interactively coexist; and the upper case is where the multimedia applications are affected without consideration of the TCP traffic performance. We see that if the multimedia applications compete with traditional application then our recommended total load for the multimedia application should be no more than 30% (1-second interval) of a given Ethernet segment. We found the average increase in delay for the traditional applications is less than 75 %, if the multimedia traffic is less than 30% of the total bandwidth available on the Ethernet. If we ignore the TCP degradation, where we have to support ONLY the multimedia applications with a reasonable quality of service on a given LAN, we could go up to 70% (1-second interval) usage of the LAN without noticeable quality of service impact. Therefore, the number of unique sessions of a given multimedia application or mix of applications should be calculated keeping this in mind for reasonable quality guarantee. Furthermore, as we increase the number of multimedia application sessions, the visual quality and delay jitter should also remain within a service tolerable range.

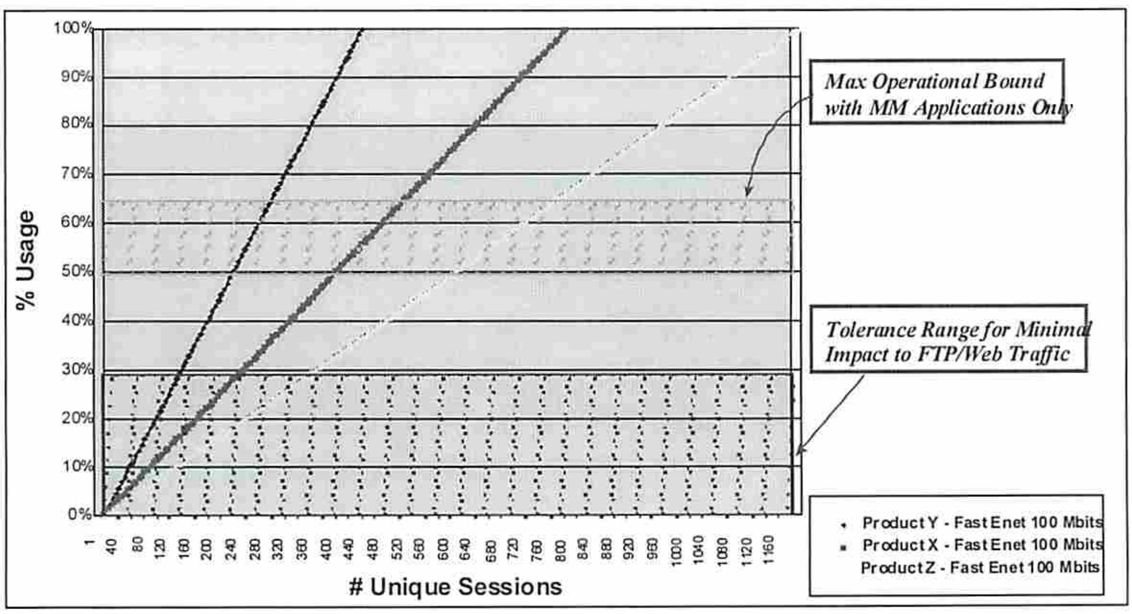
If Figure [15], we extrapolate these results to 100 Mbps Fast Ethernet and depict the same scaling profile. We base this projection on the results presented in article [1]. There are several factors involved in extrapolating the results for 100 Mbps Ethernet. Some work its advantage, e.g., 10x smaller slot size as compared to 10 Mbps Ethernet, and some to its disadvantage, e.g., the size of *"a"* (i.e.,T_p/T_x) which may make the CSMA/CD protocol performance degrade.

Figure 14: # Unique (PRODUCT X, PRODUCT Y, PRODUCT Z) Sessions v. Ethernet LAN (10Mbps)



In either case, the network planner or analyst who is looking to extend an existing or design a new network to support multimedia applications should consider several key factors in addition to the typical bandwidth sizing exercise. We suggest understanding the a) mix of traditional and multimedia applications, b) multimedia QoS and traditional applications' service level (delay tolerance) impacts, and c) multicast or unicast unique session counts. Moreover, the network protocol overhead is a factor; where we must observe the general collision overhead in CSMA/CD in our operating "ranges", for WAN serial links this overhead does not exist, and we can adjust the recommended scaling range higher, possibly 80% in the isolated case and up to 40% in the mixed traffic case. Delay jitter will play a key role in the WAN or extended LAN situations where the multimedia traffic (specifically, voice component) may be impacted more severely. More work is in progress to understand these implications in addressing network scale.

Figure 15: # Unique (PRODUCT X, Y, Z) Sessions v. Fast Ethernet LAN (100Mbps)



7. Summary

In this paper, we presented a network characterization study and corresponding results for a set of multimedia applications. We provided a practical investigation of the behavior of these multimedia applications against FTP and Web traffic, two dominant user application/services in today's corporate transaction environment. We showed through traffic, trace-driven analysis, evidence and root causes for their interactive impacts. Our key findings are that the FTP and Web have little impact to the multimedia applications, but the reverse is not true. Through our presentation, we show that the 'hard' behavior evidenced in the multimedia applications has a direct effect on the delivery response of traditional (TCP/IP) applications.

Utilizing autocorrelation and CDF analysis on the trace data, we show a practical framework for characterizing multimedia applications for network scaling and impact analysis. The main suggestions from these results direct us to the knee point for each of applications from the scaling point of view; above this point, the impact of the multimedia application on the traditional application can be severe if not planned and deployed appropriately. This suggests the need for an admission control policy to help prevent network congestion or blackout and existing time-critical business applications from suffering.

Although the scope of our study is mainly on the single segment Ethernet, this is a best-case "baseline" providing a measuring point for comparison of more complex, higher bandwidth and differing network technologies. We suggest further scaling work with multicast multimedia applications to understand the impacts of switching, router topologies, the router state space and processing time. It has been found [6] that the overhead of multicast traffic reporting (membership, joining and leaving) on a given leaf segment LAN is insignificant compared to the bandwidth available. Nevertheless, in the case of the extended LAN Backbone, where there are many LAN segments attached, the traffic overhead is not insignificant and thus should be studied further. Additionally, the probability of damage, packet disordering, loss and duplication of multicast packets in a LAN is very low, but in an internetwork, successful multicast packet delivery will decrease as the distance between the sender and group members increase.

Finally, we have some evidence of the deployment opportunities of multimedia applications today. Appropriate network characterization, planning and a more thorough understanding of the current application environment and service-levels is necessary in avoiding service impacts or costly, over-provisioned network infrastructure.

References

[1] "Evaluation of 10Base-T and 100Base-T Ethernet Carrying Video , Audio and Data Traffic" @ IEEE INFOCOM June 94 .

[2] F.A. Tobagi and N Gonzalez-Cawley, "On CSMA-CD local networks and voice communication" in proceeding of the conference on computer communication (IEEE Infocom) , March /April 1982.

[3] K. M. Nichols, "Network Performance of Packet Video on a Local Area Network," in IEEE 11th Annual International phoenix Conference on Computers and Communications, (Scottsdale, Arizona), IEEE April 1-3 1992.

[4] Abdelnaser Adas Georgia Institute of Technology, "Traffic Models in Broadband Networks" IEEE Communication Magazine .July 1997.

[5] D. Lucantoni, M.Neuts and A.Reibman."Methods for Performance Evaluation of VBR Video Traffic Models." IEEE /ACM Trans .Networking, vol.2 April 1994.

[6] Stephen E. Deering, "Multicast Routings in Internetworks and Extended LANs", ACM, 1988, pp. 55-64.

[7] Stephen Deering, Deborah L. Estrin, Dino Farinacci, Van Jacobson, Ching-Gung Liu, Liming Wei, "The PIM Architecture for Wide-Area Multicast Routing", IEEE/ACM Transactions on Networking, Vol.4, No.2, April 1996.

[8] Scott Shenker, "Fundamental Design Issues for the Future Internet."

[9] D. D. Clark, S. Shenker, and L. Zhang, "Supporting Real-Time Applications in an Integrated Services Packet Network: Architecture and Mechanism", {\it ACM} 1992, pp. 14-26.