

Performance measurements in local and wide area ATM networks

R. Eberhardt, C. Rueß, R. Sigle

Daimler-Benz AG, Research and Technology

P.O. Box 2360, 89013 Ulm, Germany

Tel: ++49-731-505-2103, Fax: ++49-731-505-4110

{eberhardt, ruess, sigle}@dbag.ulm.DaimlerBenz.COM

Abstract

Daimler-Benz Research and the International Computer Science Institute (ICSI) participated in one of the first transatlantic Asynchronous Transfer Mode (ATM) field trials. We describe the intention of the project Multimedia Applications on Intercontinental Highways (MAY), its infrastructure and results of various performance evaluations based on Classical IP over ATM (CLIP) and LAN Emulation (LANE). In addition the commonly used videoconferencing tool "vic" was evaluated in various measurements. We discuss the state of the art of wide area ATM networks and their limitations. Finally, solutions to some of the problems are outlined.

Keywords

ATM, ATM LANs/WANs, LAN Emulation, IP over ATM, Field Trial

1 INTRODUCTION

Today we can observe the evolution from export oriented to global operating companies. In order to open up new markets their special needs have to be taken into account during the development of products. This requires besides global production worldwide research and development. Integrated solutions including the areas Computer Supported Cooperative Work, Concurrent Engineering and High Performance Communications are needed to support the collaboration of engineers

spread all over the world. Therefore more flexible network infrastructures providing higher bandwidth than nowadays corporate networks are demanded.

ATM is the first networking technology which allows high speed communication for both local and wide area networks using the same network infrastructure. It is ideally suited for all kinds of applications, contrasting today's situation where different services use different network technologies. ATM has the potential to embrace both corporate and public networks, voice and data communication.

The project MAY examined if ATM can satisfy these requirements. The MAY project was initiated by T-Berkom, a subsidiary of Deutsche Telekom. The global ATM network for the MAY testbed was provided by Deutsche Telekom, Sprint, GlobalOne and Teleglobe.

This infrastructure was used by several companies and research organizations to test and develop new applications and communication protocols. Several applications in the areas cooperative engineering, distance learning, and telemedicine were envisaged. In section 2 we describe the MAY network infrastructure. We then present in sections 3 network and in section 4 application performance measurements. Finally we draw several conclusions.

2 MAY NETWORK INFRASTRUCTURE

The connection between ICSI Berkeley (California, USA) and Daimler-Benz Corporate Research Ulm (Germany) was composed of several parts made available by the service providers mentioned in Figure 1.

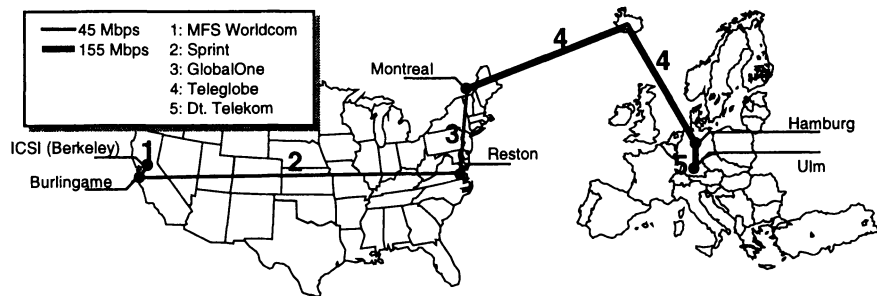


Figure 1: ATM Connection between ICSI (Berkeley, USA) and Daimler-Benz Research (Ulm, Germany)

Due to the lack of standards and implementations of B-ICI (Broadband Intercarrier Interface), PNNI-1 (Private Network Node Interface) and Q.2931 no signalling is available in current wide area ATM networks. The sole possibility to interconnect local ATM networks through wide area ATM networks is to use Virtual Paths (VP). The end systems (PCs and workstations) in both locations were connected via local ATM switches to the wide area ATM network.

The network providers Sprint, GlobalOne and Teleglobe used spare capacity on their ATM links. The VP from Berkeley to Hamburg was permanently configured. In order to setup an end to end ATM connection a VP from Ulm to Hamburg had to be reserved and configured by Deutsche Telekom's ATM network management center in Cologne. The local ATM switches and end systems had to be configured on both sites.

To be able to reuse existing software without any changes we used Classical IP over ATM [RFC1483, RFC1577] and LAN-Emulation Version 1.0. Because the network providers only offered a constant bit rate (CBR) service, we had to use traffic shaping in the end system ATM adapters.

3 NETWORK PERFORMANCE MEASUREMENTS

In order to examine which kind of applications can be run on top of the transatlantic testbed we made several performance measurements. We tested the throughput and round-trip delay of TCP and UDP using CLIP and LANE in the local and transatlantic ATM network. For the transatlantic connection an extended TCP implementation [RFC1323] was used which solved the problem of "Long Fat Networks (LFN)".

3.1 Benchmark programs and utilities

The public domain tool "netperf" [np] from Hewlett-Packard allows several kinds of performance measurements. Of special interest are measurements of the bulk data transfer performance and request/response delay of network connections using TCP and UDP.

To measure the throughput netperf enables the variation of socket and message sizes on the sender and receiver side. In the case of datagram oriented connectionless transport protocols like UDP the loss rate is of special significance. All measurements were performed over a period of 60 seconds.

To measure the CPU load of the end systems, the tool "sar" was used which is included in the Solaris operating system. The CPU load was measured every five seconds during the sending/receiving period.

To get information about the AAL5 PDUs sent/received by the Fore ATM adapter, the tool "atmstat" was used. Atmstat is included in the Fore driver software.

3.2 Network configuration

As end systems several SUN workstations with varying performance were used: From Sparc 5 to Sparc Ultra 2 with dual CPU. All the workstations have a FORE SBA-200 ATM adapter. Tests were performed running Solaris 2.5 and 2.6 beta operating systems. The ATM Switch in Ulm is a Fore ASX 200 and in Berkeley a Bay Networks LattisCell switch. Figure 2 shows the test configuration.

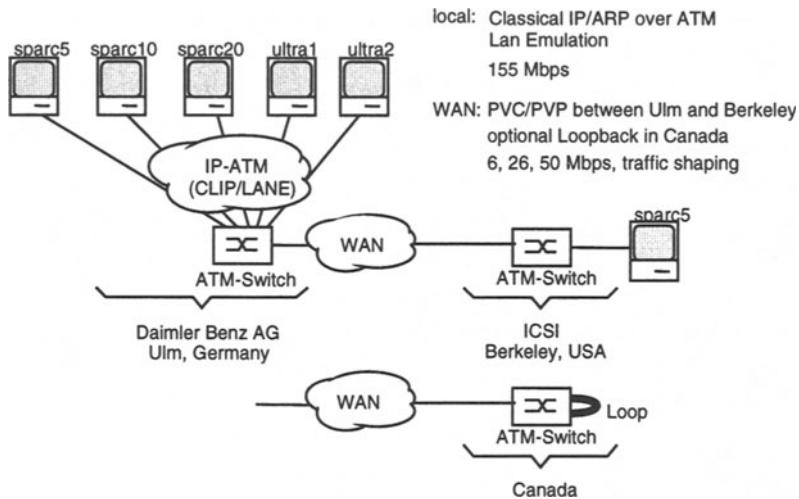


Figure 2: ATM test environment

The workstations use CLIP or LANE. In the wide area network permanent virtual connections (PVCs) or Permanent Virtual Paths (PVPs) were established. Optionally a loop was configured on a switch in Canada connecting hosts in Germany with about the same delay. The PVCs were configured with a bandwidth of 6, 26 and 50 Mbps. Using the loop over Canada also simplified the task of configuring the end systems, as they were at the same location. The transmitting end systems used traffic shaping to avoid cell loss in the interim switches of the wide area network providers.

3.3 Performance in local ATM networks

This chapter summarizes the results of the performance measurements in the LAN. First the TCP performance of Classical IP over ATM was evaluated. In a second step TCP using LAN Emulation was considered. The chapter closes with some UDP measurements using CLIP.

TCP performance using Classical IP over ATM

The TCP throughput using Classical IP was measured between the Sun workstations sparc10 and sparc20 with varying message sizes. The buffer sizes of both sockets were set to 64 kByte.

The message size was increased in steps of 1024 Bytes. During the measurement the workstations and the ATM network weren't used otherwise.

Due to the lack of the computing power the workstations couldn't make use of the full bandwidth provided by the ATM adapters (Figure 3a). Without the overhead of several protocol layers (SONET-OC3, ATM, AAL5), AAL5 can provide a maximum bandwidth of 135.6 Mbps [atmarp], [krivda]. CLIP and TCP overhead

are reducing this value because of SNAP header, AAL5 trailer and padding to 135.1 Mbps using the default MTU size of 9180 bytes.

The ATM adapters have an onboard RISC processor and special hardware to process the ATM cells. They are providing an interface to the ATM adaptation layer 5 (AAL5). Therefore the processing of ATM cells doesn't add further load to the host computers.

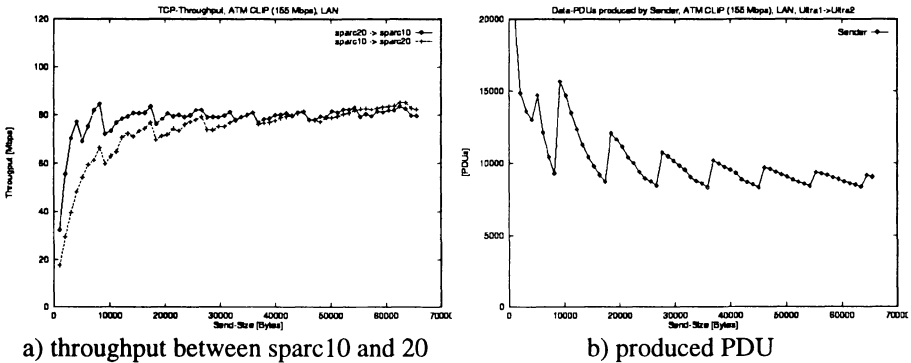


Figure 3: TCP measurements in a LAN environment

The transfer rate decreases regularly about every 8 kByte. This is due to the MTU size of 9180 bytes in CLIP. When a TCP packet larger than the MTU size is transmitted it has to be fragmented at the sender side and reassembled at the receiver side. This operation is done within the processor of the end system and decreases the overall performance. With bad fragmentation (e.g. 9216 or 18432 bytes) the protocol overhead increases additionally [mol].

Figure 3b shows, how the number of PDU increases when the packet size skips above a multiple of the MTU-size. The size of one PDU is limited by the MTU-size. Therefore an additional PDU is necessary each time the packet size steps over a multiple of the MTU-size. Figure 3b relies on the results of the measurements with Ultra Sparc workstations.

These workstations seem to have sufficient CPU performance to make use of the available bandwidth. A throughput of up to 130 Mbps was achieved.

The CPU load during this test was about 20% for the Ultra 2 with dual CPU and 40% for the Ultra. For small packet sizes the CPU load is significantly higher, e.g. for a packet size of 1024 Byte the CPU of the Ultra 1 was at its limits (100%).

To get a direct comparison of the CPU load of the sender and the receiver, a TCP transmission to "localhost" was measured. The tool "sar" only shows the total CPU load of a host. To get a process related value for the CPU load, we used the publicly available tool "top". Figure 4 shows the result of the measurement for three different Send-Sizes.

For very small packet sizes, the sending process needs more CPU power than the receiver. For large packet sizes, the load gets more and more equalized.

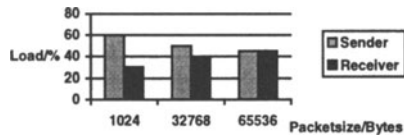


Figure 4: CPU load of sender and receiver (TCP)

TCP performance using LAN Emulation

For LAN Emulation similar measurements were made. In comparison to Classical IP, LAN Emulation adds an additional protocol layer and thus additional overhead to the communication. The default MTU size for (Ethernet-) LAN Emulation is 1500 bytes.

The throughput of LANE for a TCP stream between Sun Sparc Ultra computers reaches about 110 Mbps and is close to the theoretical upper limit. For small packet sizes the throughput decreases to 60 Mbps (1024 bytes packet size).

The CPU load for these measurements is in the range of 80-100% (Ultra 1) depending on the packet sizes. Compared to the CLIP results, the load is about twice as high.

The number of PDUs generated with LANE is about five times (50000) as high as with CLIP. This is due to the reduced MTU size for LANE

UDP measurements using CLIP

The UDP measurements show that the maximum transfer rate is limited by the processing power of the end systems. Due to the lack of flow control in the UDP protocol a slow receiver can be overloaded by a fast sender.

In our test case we send 62 Mbps from a SUN SPARCstation 10 to a SUN SPARCstation 20. In the other direction the SPARCstation 20 could send 102 Mbps but the slower receiving workstation wasn't able to process this overload leading to a packet loss rate of almost 100 percent. Similar results are expected for UDP on top of LANE (not measured).

3.4 Performance in a transatlantic ATM network

For the WAN measurements, Classical IP was used. Most of the examined effects correspond to the TCP protocol. LAN Emulation would therefore make no difference. The following configurations were setup:

First we tested a default implementation of TCP between two hosts in Ulm and Berkeley. The available bandwidth for this connection was set to 6 and 26 Mbps; the roundtriptime was about 190 ms.

In a second test, we made similar measurements with an extended implementation of TCP supporting RFC1323. These tests were performed between hosts in Ulm using the loopback in Canada. The roundtriptime of this connection was also about 190 ms. In order to measure the performance in the wide area network a PVC was

setup between two workstations (Figure 2) and cell rates of 6, 26 and 50 Mbps were used.

To avoid cell loss in the ATM switches in the wide area network the traffic was shaped on the sender side to the peak cell rate [atmarp].

TCP performance with standard TCP implementation

For the TCP tests the message sizes were varied and socket sizes were set to the maximum value of 64 kByte. With the 6 Mbps as well as with the 26 Mbps connection only a fraction of the provided bandwidth could be used (approximately 2.2 Mbps). The increase in bandwidth from 6 to 26 Mbps had almost no effect to the results.

Sun's standard TCP implementation in Solaris 2.5 cannot exploit the available bandwidth in an environment with a high-speed network connection and large round-trip delay. The problem of a high path capacity¹ is already wellknown. A sender of TCP packets has to wait until the transmitted packets are acknowledged by the receiver (sliding window protocol [comer, brzi96]). Assuming the sender waits for an acknowledgment after it has sent the maximum window size of 64 kByte means that a maximum of 64 kByte could be on the wire. With a round-trip delay of 190 msec the maximum throughput is $64 \text{ kByte} / 190 \text{ msec} \approx 2.7 \text{ Mbps}$. Several TCP extensions have been suggested to solve this problem [RFC 1323], some are available in beta implementations (see below).

TCP with extensions for Long Fat Networks (LFN)

RFC1323 suggests an extension to the TCP protocol to support window sizes greater than 64 kByte. An implementation of this extended TCP version should be capable of using the bandwidth of ATM connections with high delay (Long Fat Networks). A beta release version of the Solaris operating system (2.6) which implements the extensions of RFC1323 was used during these tests.

To enable the operating system to use the "Big Window" option, a couple of "nnd" commands have to be issued. Ndd is a Solaris command to manipulate driver parameters. In particular the buffer space for the socket can be adapted according to the bandwidth*delay product of the connection

Table 1 shows the results of the WAN measurements. For the 6 Mbps connection the roundtrip time was approximately 190 ms, which leads to a bandwidth*delay product of about 140 kByte. After adapting the TCP parameters with the "nnd" program, we achieved a throughput of 5.05 Mbps. An increment of the maximum window size to 1 MByte did not make any difference. For this data rate the Send-Size did not affect the throughput.

For 50 MBps the bandwidth*delay product is about 1.2 MByte. With buffers of 1.5 MByte we could only use 60% of the available bandwidth. To investigate this

¹ bandwidth and roundtrip delay product

problem, measurements in the LAN environment were carried out, especially concerning the traffic shaping capability of the Fore adapter.

Table 1: WAN measurements with TCP supporting RFC1323

Avail. BW	buffer	send size	Throughput	CPU send	CPU rcv
6 Mbps	140000	1k-64k	5.04 Mbps	4%-1%	1%
50 Mbps	1572864	1024	31.45 Mbps	28%	5%
		32k, 64k	~29.83 Mbps	~13%	3%
		1024	21.35 Mbps	18%	4%
25 Mbps	1048576	32k	19.17 Mbps	6%	2%
		1024	18.29 Mbps	17%	5%
		32k, 64k	~17.64 Mbps	~8%	~2%

In the LAN environment the traffic shaping of the Fore adapter was set to different values using the "atmarp" command. The possible throughput was measured for a send size of 32768 bytes. The results are shown in Figure 5. The theoretical boundary is also given. The traffic shaping parameter of "atmarp" corresponds to the 48 byte payload of the ATM cells. As the figure shows, the characteristics of the traffic shaper is very inefficient for bandwidths greater than 40 Mbps.

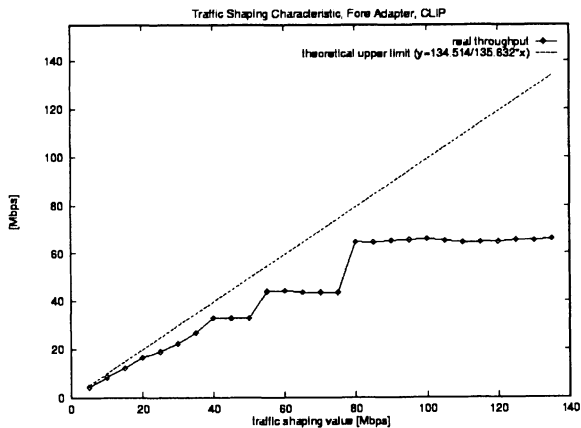


Figure 5: Traffic shaping characteristic of Fore adapter

You can also recognize that the WAN results (Table 1) are as good as the results from the LAN case. The problem of the bad throughput results from the traffic shaper in the workstation and not from the TCP implementation.

UDP results in WAN environment

For an available bandwidth of 6 Mbps there were no packet losses.

The 26 Mbps measurements were carried out from a SUN SPARCstation 20 to a SUN SPARCstation 5. The sending computer was able to send with the full

bandwidth. Only for small packet sizes losses could be observed at the receiver side. In this test case the ATM connection was monitored to guarantee that these errors weren't caused by cell losses of the ATM connection.

Rather it looks like the receiving workstation was already at its limits. In this case the receiver is much more loaded due to the processing of the UDP packets. Additionally every received packet releases an interrupt at the receiver side.

4 PERFORMANCE OF THE VIC APPLICATION

The MBone tool "vic" [vic] is widely used for video conferencing [MBoneApps, kumar]. In this section the measurements concerning the tool vic are presented.

The MBone tools support a wide range of compression schemes and workstations. If the workstation doesn't have a special hardware support for the used compression scheme most of the operations have to be done in software. Especially for video, this causes a nonnegligible CPU consumption. Therefore we carried out several measurements in order to evaluate the bandwidth and computation requirements using the vic application.

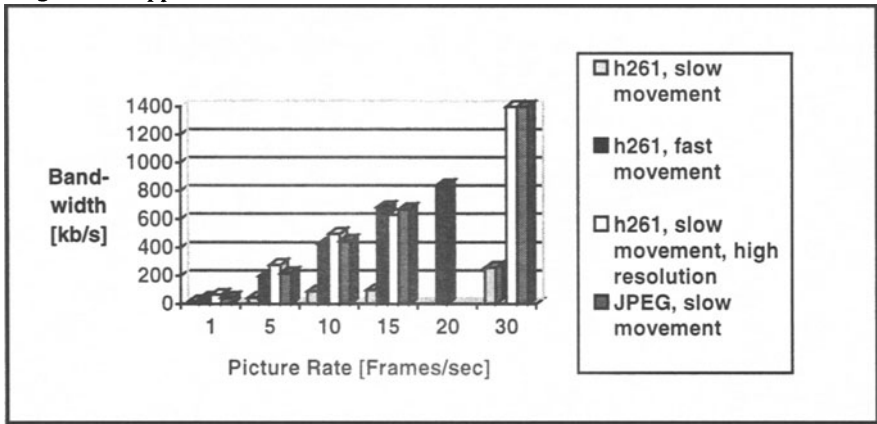


Figure 6: Bandwidth Requirements

For our tests we used two Ultra Sparc high performance workstations. To separate the effects caused by encoding and decoding of the video we first used unidirectional video transmission. We used h261 and JPEG video encoding (CIF format) and run tests with slow and fast moving people in front of the camera (slow and fast changing scene). We also run a test with high resolution video transmission. In these tests we varied the picture rate between one and thirty frames per second. We evaluated the bandwidth requirements using the measurement embedded in vic. Figure 6 shows the results of these tests.

The needed bandwidth increases linearly with the picture rate, as expected. As h261 uses an inter picture encoding mechanisms, the bandwidth requirements for slow changing scenes of h261 are much lower than this of JPEG. It can be seen, that the

bandwidth requirements of h261 increases significantly when the movement of the people in front of the camera becomes very fast. Moreover, due to end system limitations we couldn't send video at 30 frames per second using h261 encoding in the fast moving case (maximum picture rate: 20 frames per second). As the results from CPU measurements show, this limitation was caused by I/O operations the application has to wait for. The bandwidth requirements of high resolution h261 encoded video for slow changing scenes is similar to the requirements of JPEG with medium quality.

Parallel to the bandwidth measurements we observed the CPU consumption on the sender and receiver side using the sar tool (Figure 7 and Figure 8, respectively).

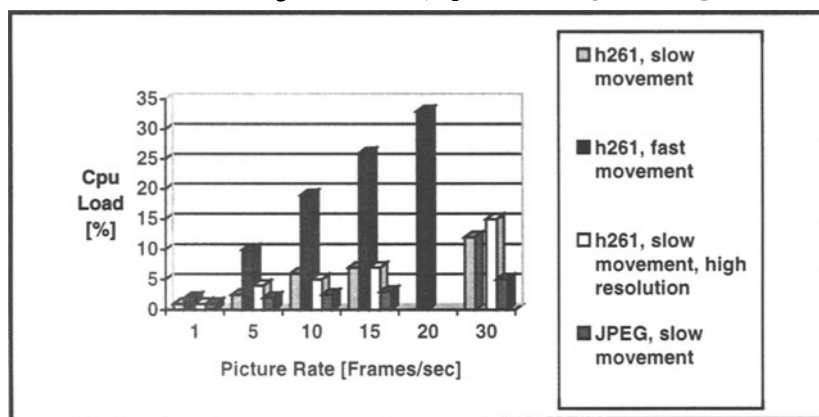


Figure 7: CPU load at the Sender

Like the bandwidth the CPU load increases almost linearly with the picture rate. The measurements show that the sender's CPU is much more loaded when using h261 than using JPEG. In the slow movement case the CPU load at the receiver is smaller for h261 than for JPEG, but this changes in the fast movement case. The figures also show that with h261 faster changes in the scene have stronger influences on the CPU load than changing the picture resolution.

Our bidirectional tests (results not presented here) show, that the CPU load of sender and receiver of the unidirectional case must simply be added to get the figures for the bi-directional case. Our measurements show, that software encoding and decoding cause significant CPU load which can't be used for other applications (e.g. application sharing) running in parallel. Videoconferences with more than two participants or between lower performing workstations increase the problem further. Nevertheless the Mbone tools are applicable in cases where video is just used in the starting phase of a conference with little number of participants (i.e. for greeting purposes) when no other application is running. They also satisfy the needs when only small picture rates are requested.

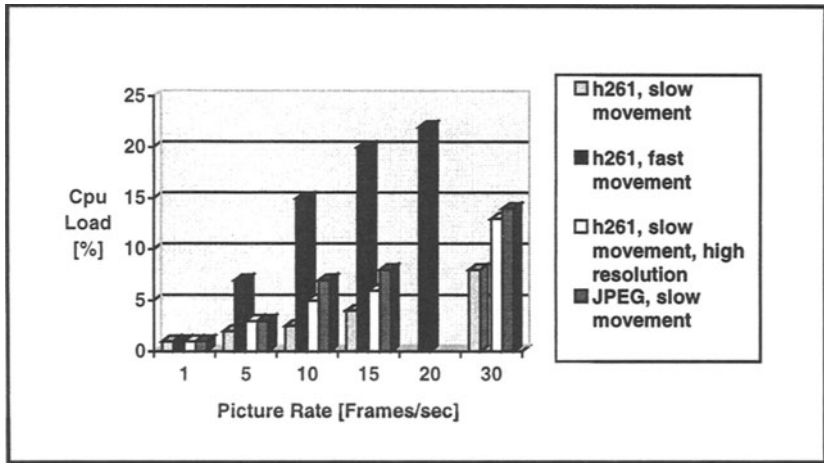


Figure 8:CPU load at the Receiver

To use the end system and network resources in an optimal way multimedia applications need mechanisms to throttle down the sender's transmission rate if the receiving end system is already saturated or if the network can't guarantee the requested quality of service anymore [AlSi96].

5 CONCLUSIONS AND FUTURE WORK

Our experiences with ATM in global networks show that up to now ATM's potentials are only used to a limited extend. To exploit them, enhancements in the provided services and protocols are needed. Especially signalling in local and wide area networks and means to monitor connections must be made available by the manufacturers and network providers. New lightweight protocols that take advantage of the ATM QoS guarantees must be developed. Further research for efficient multicasting over ATM is required. Also transport layer protocols need to be adapted to environments with high bandwidth delay products. In order to support worldwide collaboration of people new applications must be developed or adapted to a global high speed network environment [EbRuSi97]. Future work is dealing with the provision of QoS to the applications. This can be done using for example the WinSock 2 API or by using RSVP in conjunction with ATM [BrSt97].

6 REFERENCES

- [AlSi96] Alfano M., Sigle R., *Controlling Resources in a Collaborative Multimedia Environment*, Proceedings 5th IEEE International Symposium on High-Performance Distributed Computing (HPDC-5), Syracuse, USA, 1996
- [atmarp] Manual page atmarp, Fore Systems, 2/1996
- [BrSt97] Braun T., Stüttgen H.J., *Implementation of an Internet Video Conferencing Application over ATM*, IEEE ATM'97 Workshop, Lisbon, Portugal, 1997
- [brzi96] Braun T., Zitterbart M., *Hochleistungskommunikation, Band 2: Transportdienste und -protokolle*, Oldenbourg 1996, ISBN 3-486-23088-3
- [comer] Comer, Douglas E., *Internetworking with TCP/IP*, Volume I, Second Edition, Prentice Hall 1991, ISBN 0-13-468505-9
- [EbRuSi97] Eberhardt R., Rueß C., Sigle R., *Multimediale Anwendungen in globalen ATM-Netzen*, Proceedings of GI/ITG Kommunikation in Verteilten Systemen (KiVS'97), Braunschweig, Germany, 1997
- [krivda] Krivda, Cheryl D., *Analyzing ATM Adapter Performance, The Real World Meaning of Benchmarks*, Efficient Networks Inc., 1996, <http://www.efficient.com/doc/EM.html>
- [kumar] Kumar V., *MBone: Interactive Multimedia On The Internet*, Macmillan Publishing, Simon & Schuster 1995
- [MBoneApps] Kumar V., *Mbone Desktop Applications*, <http://www.best.com/~prince/techinfo/mc-soft.html>
- [mol] Moldeklev, Kjersti., *The effect of end system hardware and software on TCP /IP throughput performance over a local ATM network*, Teletronik, 1995
- [np] Netperf, www, <http://www.cup.hp.com/netperf/NetperfPage.html>
- [RFC1323] Jacobson V., Braden B., Borman D.: *TCP Extensions for high-performance*, Request for Comments 1323, May 1992
- [RFC1483] Heinanen J., *Multiprotocol Encapsulation over ATM Adaptation Layer 5*, Request for Comment 1483, Juli 1993
- [RFC1577] Laubach, M., *Classical IP and ARP over ATM*, Request for