Scalability in Analyzing the Availability of Large-Scale Networks using Multicast

Falko Dressler*
University of Erlangen-Nuremberg, Nuremberg, GERMANY

Abstract

The goal of this paper is to identify and discuss scalability issues for the measurement-based analysis of the availability in large-scale networks using IP multicast technology. Additionally, a solution for this task is proposed named MQM (multicast quality monitor). The utilization of IP multicast in economically critical environments such as the distribution of stock exchange prices is increasing. Therefore, the availability requirements on this service are increasing as well. Unfortunately, no measurement tools exist which determine the global multicast reachability to provide the required availability information. A novel solution is discussed in this paper, which claims to provide a high scalable measurement environment. The behavior and performance of the proposed approach was analyzed using a simulation model as well as using lab measurements.

Key Words: Network analysis, multicast, availability control, network monitoring, distributed systems.

1 Introduction

The increasing employment of IP multicast even for mission critical applications such as the distribution of stock exchange prices in the stock market [9] has motivated this work. To provide such a high availability as required for such applications, reachability measurements have to be employed as discussed in [14, 15]. The analysis of the availability is demanded by all serious users of the global IP multicast network. Even if IP multicast helps to save resources for a one-to-many transmission over the internet [10], there are a few drawbacks hindering multicast to get employed for mission critical applications. One problem in the current global multicast infrastructure is the absence of adequate measurement tools. Even if there are first approaches to test the functionality of the multicast network since the early beginnings of the development of IP multicast routing protocols, all these concepts do not allow a scalable deployment over large parts of the network or even in the global multicast enabled internet [2].

Typical problems of testing the availability of multicast networks [5], i.e., scalability and completeness, are introduced in

this paper accompanied by a short overview to the state of the art in availability measurements in multicast environments. Additionally, a new approach is presented, the multicast quality monitor (MQM) [8]. The concepts of this idea allow a high scalable availability analysis even in large scale multicast networks [7].

The main contributions of this paper are to provide a detailed view to issues in quality of service measurements in IP multicast networks and the presentation and discussion of the multicast quality monitor. This method was tested in real life scenarios as well as in simulation experiments.

The rest of the paper is organized as follows. Section 2, summarizes the most prominent issues in IP multicast measurements followed by a study of related work in Section 3. The primary concepts of the MQM are discussed in Section 4. Results from measurement and simulation experiments are shown in Section 5. A conclusion summarizes this paper including some outlook to other ongoing work.

2 IP Multicast Measurement Issues

Several issues have to be addressed for successful utilization of IP multicast services. Besides networking objectives such as to provide a minimum (or even a guaranteed) amount of quality [17], the monitoring and analysis of the current behavior of IP multicast networks must be considered [11, 16]. Two of the most important issues: scalability and completeness of the measurement are described in the following.

2.1 Scalability Issues

Scalability is always an issue in multicast environments due to the working principles of multicast routing [17, 19]. In the context of this work we measure scalability in terms of the number of messages needed for one measurement compared to the number of participating multicast nodes. Ideally, a fixed number of messages is needed. The basic concepts of multicast are as follows: the sender of a packet stream sends its packets only once to a so-called multicast address and the network is responsible to deliver the message to each client who is interested in receiving traffic for this particular multicast address. Therefore, each of these clients receives a copy of each packet sent to the multicast group. The scalability of an availability test strongly depends on the concept of the message passing between

---
* Autonomic Networking Group, Dept. of Computer Science 7. E-mail: dressler@ieee.org.

all test-stations that will be called probes in the following. The easiest approach is to have each participating probe sending test packets on a regularly basis. These test packets might be responded by sending answer packets back to the originator or – which allows for more precise measurements – to the multicast group allowing each other probe to receive this response.

An example for such a measurement is the multicast beacon, which is described later in the related work section. The scalability of this approach is shown in Figure 1. Depicted is the
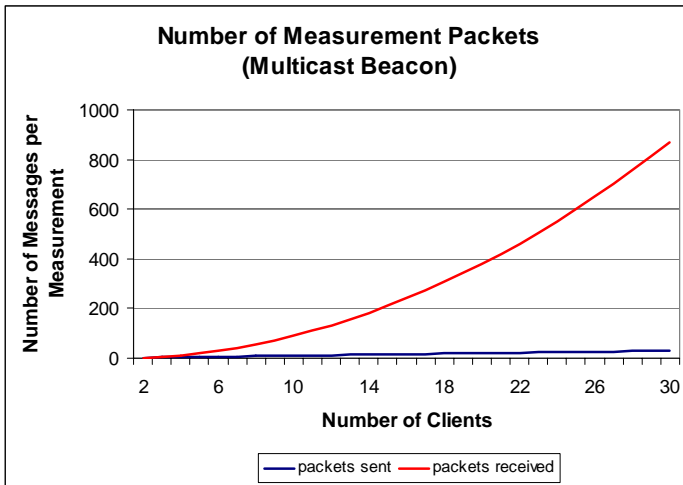

Figure 1: Scalability of a reachability test using the multicast beacon

relationship between an increasing number of clients and the resulting number of messages per measurement. It can be seen that the number of messages that have to be received and analyzed at each participating client increases dramatically according to the formula $n = p * (p - 1)$ where p is the number of involved probes. The amount of wasted network utilization b can be calculated as follows:

$$b_{stream} = \text{packet rate} * \text{packet size}$$

$$b_{total} = b_{stream} * n.$$

For example, the multicast beacon sends about 10 packets per second with a size of 100 bytes each. Thus, the resulting bandwidth requirements b_{total} at each probe for 30 participating clients is about 7 Mbit/s which is too much just for measurement traffic besides the regular network usage.

2.2 Completeness of Availability Analysis

Besides the scalability, there is an issue of importance depending on the measurement concept as well: the completeness of the analysis. In case of failures and partially unavailable network parts, i.e., split sub-networks, the knowledge about the internal behavior of each of these parts is required in order to provide a complete analysis. Due to the high complexity of multicast routing protocols and due to incomplete

and non-interoperable implementations, network partitioning appears to be very common in well-known multicast networks [13]. The problem is described in more detail in Figure 2. The functional multicast network (a) can be split into several partitions (b) due to a failure in multicast forwarding at some point in the network. It is up to the measurement system to analyze this behavior and to provide information about each particular sub-network.
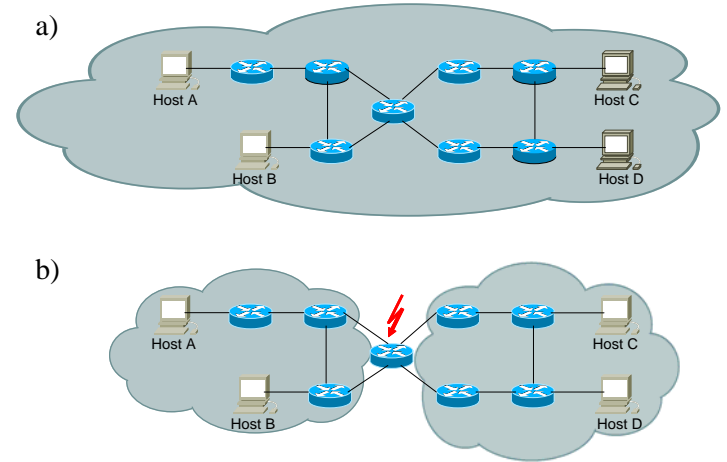

Figure 2: Network partitioning

Typically, it is not possible to get information about a (sub) network if there is no measurement probe deployed in it that is actually sending test packets. Therefore, most approaches are based on the concept of building probes that periodically send data packets and reply on received requests. Thus, the scalability issues discussed in advance apply again. New concepts have to be developed which allow a high scalability accompanied by a complete availability analysis. A possible solution is discussed in Section 4.

3 Related Work

There are two compatible approaches to the here presented novel solution, namely the multicast reachability monitor introduced by Almeroth [3] and the multicast beacon [4] from the NLANR (National Laboratory for Applied Network Research). Both are described and analyzed in the following.

3.1 Multicast Reachability Monitor

The multicast reachability monitor (MRM), formerly known as the multicast route monitor was developed to allow a centralized reachability measurement based on probes located all over the multicast network. End systems can be used as probes as well as the multicast routers themselves. The MRM, which started as an IETF draft [1], defines three different processes: the MRM manager, the test sender, and the test receiver. Controlled by the manager, the multicast reachability monitor is able to create a configurable packet flow at each test sender. Using the received packets, the test receivers are able to compute measurement results, such as the packet loss ratio, which provide

a good estimation of the reliability of the network.

Comparing the abilities of the multicast reachability monitor with the described problems or more precisely the scalability and the completeness, it has been shown that the deployment of the MRM can be either in some degree scalable, i.e., if only one or a few senders are implemented. Unfortunately, network partitioning cannot be recognized in this case and the connectivity can only be tested in one way from the sender toward the receivers. On the other hand, the MRM has strong scalability problems if all the receivers are working as senders as well. In this case, the network utilization for the measurement is much too high.

3.2 Multicast Beacon

The multicast beacon is the result of a research project from the NLANR. Currently, there is an implementation in JAVA for the so-called beacon clients available, which should run on nearly every end system with an installed JVM (java virtual machine). The so-called beacon server consists of a perl program. The principles of the multicast beacon and the MRM are very similar. The definition of the multicast beacon includes a server computing the QoS parameters from measurement results and the clients, named beacons, which are sending and receiving the measurement packets. All the beacons interact directly with each other by constantly sending IP multicast packets to an administratively configured multicast group. Each beacon client reports its measured data, i.e., the results of received packets (beacons) to the server. The server calculates a matrix including each active client and allows these results to be accessed via a web gateway.

The main differences between the MRM and the multicast beacon are the capability of the multicast beacon to provide a direct access to the measurement results and the wider range of QoS measurements (packet loss ratio, delay, and jitter). On the other hand, the MRM allows one to distinguish between a test sender and a test receiver. This differentiation results in a much lower impact on the network, especially if broadcasting scenarios are the most common applications in the particular network under study. The scalability was already discussed during the introduction. It was shown that this approach is not scalable for application in very large scale multicast networks.

3.3 Summary

As shown above, none of the available tools is able to fulfill the requirements for a complete and scalable availability analysis. A brief summary of the drawbacks is provided in

Table 1. New concepts are required. An approach to solve the problems is described in the following section.

4 Methodology: The Multicast Quality Monitor

The focus of this section is a new multicast ping mechanism introduced as part of the multicast quality monitor [6]. The primary goal of this new methodology is a high scalability. The MQM ping mechanism was designed for a complete analysis of the reachability and, therefore, of the availability of large scale multicast networks.

4.1 MQM Ping Mechanism

The MQM ping mechanism is directly based on the working principles of IP multicast. As shown in Figure 3, the MQM ping relies on the transmission of ping request messages and the reaction to the reception of such packets, the transmission of corresponding response messages. Without restricting the generality, we believe that probe P_1 is sending a MQM ping request packet. Based on the working principles of IP multicast, all the other probes receive this request and start sending a response message. Therefore, P_1 gets an answer from each participating probe and is able to analyze the reachability in the network. As shown in note 1, e.g., P_2 receives an answer from
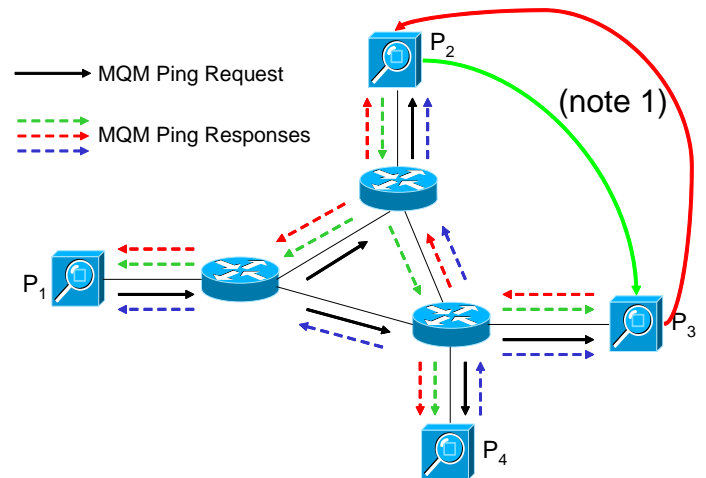

Figure 3: MQM ping mechanism. Shown are four MQM probes ($P_1 \ldots P_4$) connected by three routers. One MQM Ping Request (bold arrows) causes a number of MQM Ping Responses (dashed arrows). The emphasized connection between P_2 and P_3 (note 1) is explained in the text.

Table 1:   Summary of the drawbacks of established multicast measurement tools

	Multicast Reachability Monitor	Multicast Beacon
Scalability (smooth adaptation to a large network environment)	Depending on the number of senders	Not at all
Completeness (capability to detect network partitioning)	Not at all	Detection is possible by the server. This requires a working unicast connection to from each probe to the server.
Flexibility (possibility to measure the current network quality as well)	Limited, only unidirectionally between the probes towards the server.	Yes.

P_3 as well as vice versa. Using these receives response messages, P_2 and P_3 can analyze the behavior of the network between them. Thus, using a single ping request it is possible to analyze the complete multicast network.

In contrast to other tools such as the mentioned multicast reachability monitor or the multicast beacon, the MQM ping is not only based on a request-reply mechanism. It allows use replies only to get enough information about the connectivity in the network to provide a complete analysis of the availability.

The goal of a high scalability accompanied by a high fault tolerance is achieved by using the following concepts:

- The overall MQM system must ensure to send one MQM ping request per minute (all the typical multicast routing protocols have a timeout for the entries in their routing tables of three minutes, thus, the states must be refreshed within this interval). This MQM ping request can be sent by any of the available probes.
- For higher reliability, two active MQM ping requests must be ensured in each time interval (i.e., per minute) and per (sub) network (the messages are unreliable and, therefore, may get lost). In other words, each participating MQM probe must receive two MQM ping requests per minute.

This can be achieved in an implementation using the following methodology:

- At the startup a single MQM ping request is sent (enabling all the other probes to learn about the new participant).
- In the following, a MQM ping request is sent only if there were less than two requests received in the last interval (in the last minute). – This configuration relies on the behavior of typical multicast routing protocols: usually, the timeout for active multicast groups is between two and three minutes, therefore, at least one packet must be sent within that interval in order to refresh the routing tables. Secondly, due to the unreliable UDP-based communication, the requirement for two messages in a single interval reduces the possibility to miss too many requests.

Regarding the scalability of the MQM ping mechanism, it can be said that a proportional scaling can be achieved. A graph showing the scalability is provided in Figure 4. Obviously, the increase of required messages is no longer exponential with an increasing number of probes as shown for the multicast beacon. Now we achieved a linear scaling, which is much more feasible even for large-scale networks. We consider to call this linear increase scalable because it significantly outperforms all other solutions. The novelty of the shown ping mechanism is its ability to work in a multicast environment without the common problem of packet explosion. Based on the working principles of the new multicast ping, it is possible to detect network partitioning (in each partition must be ping requests sent by any participating probe). Additionally, there is the requirement to send further requests if some request packets got lost due to overload situations.

In summary it can be said that the MQM ping mechanism

allows a high scalable reachability measurement of IP multicast networks, i.e., it reduces the impact on the network to a very low level. For example, if 30 probes are employed, a theoretical peak in network utilization due to the overhead of the measurement of about 46 kBit/s is feasible (100 Byte per packet, 2 requests and 29 responses per minute). In practice, not all packets in each interval will be sent at the same second and, therefore, the wastage of network resources will be considerably lower.
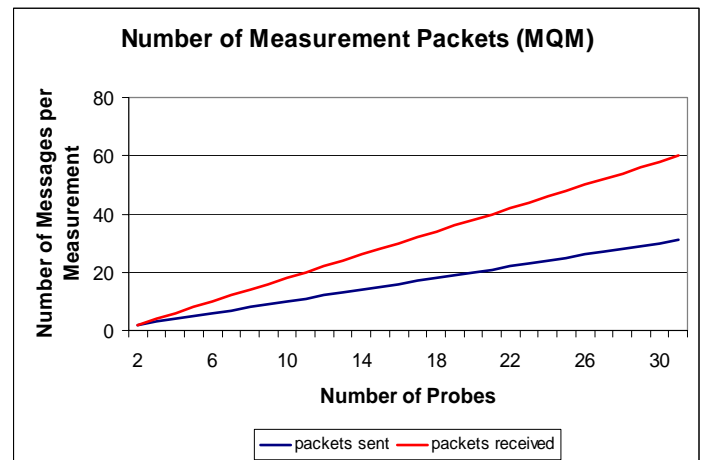

Figure 4: Scalability of the MQM ping mechanism

4.2 Availability Analysis

Normally, the reachability should be maintained by the simplest possible mechanisms. Redundancy is provided in typical backbone networks. Even if the provisioning of the connectivity sounds easy, especially in IP multicast environments, this cannot be presumed. The multicast reachability suffers from the complexity of the multicast routing protocols and the lack of experience of network administrators declining with these mechanisms. Another problem is the still miserable interoperability between devices of different manufacturers and, partially, the incomplete implementation of the protocol stacks.

Using the results of reachability measurements over a period of time, the reliability of the network can be calculated. High availability systems require a reliability of nearly 100 percent. Therefore, reachability means connectivity at a certain point of time and reliability stands for the percentage reachability over a period.

Based on the reachability measurements using the MQM, it is possible to estimate the availability of single network paths as well as of large network parts. The results of the single measurements are distributed over all the employed measurement probes. This results in a typical problem in distributed systems, how to retrieve the required information with a minimum transmission overhead. This kind of problem was discussed in many facets, particularly in peer-to-peer networks. We also thought of using IP multicast for the collection of the measurement results but decided to use reliable unicast transmissions based on results of [12] and due to the

required overhead for performing the reliable transmission.

Depending on the amount of measurement data, it seems to be advisable to preprocess the information at each probe before they are transmitted to a common analyzer. Finally, the information must be centrally collected and analyzed. For example, it is possible to conclude the availability of a network path by looking for received messages along the path. It must be considered that a received packet can only be used to determine the unidirectional connectivity between the two associated probes.

An example of such an analysis is shown in Figure 5. In this sample network, there were no correctly received measurement data packets between P_4 and P_2 as well as between P_3 and P_2. The conclusion is that there must be a failure between R_3 and R_4 or, maybe, only at R_4. If there is a recording of the network behavior available over a period, information can be gathered whether there is a temporary failure or not.
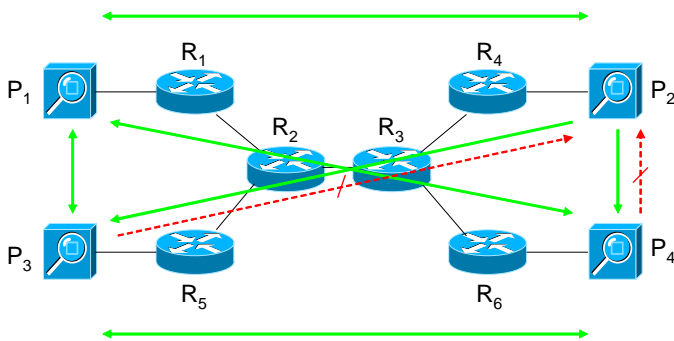

Figure 5: Availability analysis

4.3 Summary

In summary, it can be said that the analysis of the availability of an IP multicast network using the mechanisms of the multicast quality monitor is highly scalable and fault tolerant. Thus, there are mechanisms available that allow an examination of multicast networks to determine the reachability, for example, to verify a SLA (service level agreement) which guarantees some degree of availability for a particular network connection or even for a larger part of the network.

<div align="center">

5 Measurement and Simulation Results

</div>

We executed a number of lab experiments with our MQM implementation. Basically, these measurements demonstrate the capabilities of the MQM. Additionally, we created a simulation model to verify the expected behavior. The experiments were primarily conducted to show the applicability of MQM in a real network scenario. On the other hand, the simulation model was used mainly to evaluate MQM in terms of necessary messages and overhead.

5.1 Lab Experiments

Several MQM probes were used to evaluate the QoS measurement with the MQM. All four probes are hosted in distinct subnets. We configured a transmission rate of 10 packets per second. The size of each transmitted packet is 1400 bytes. The corresponding results of the one-way delay measurements are shown in Figure 6. The delay measurements within our local network (between two Ethernet segments) indicate an OWD of less than 2 ms on average. A single peak of about 16 ms was possibly caused by operating system properties. The second figure depicts a measurement to a remote probe (this probe was installed in a different lab with a diameter of about 8 hops). Obviously, the OWD is less stable. The measurements fluctuate between 20 ms and 50 ms. Additionally, peaks of about 250 ms indicate higher dynamics in the network as well.
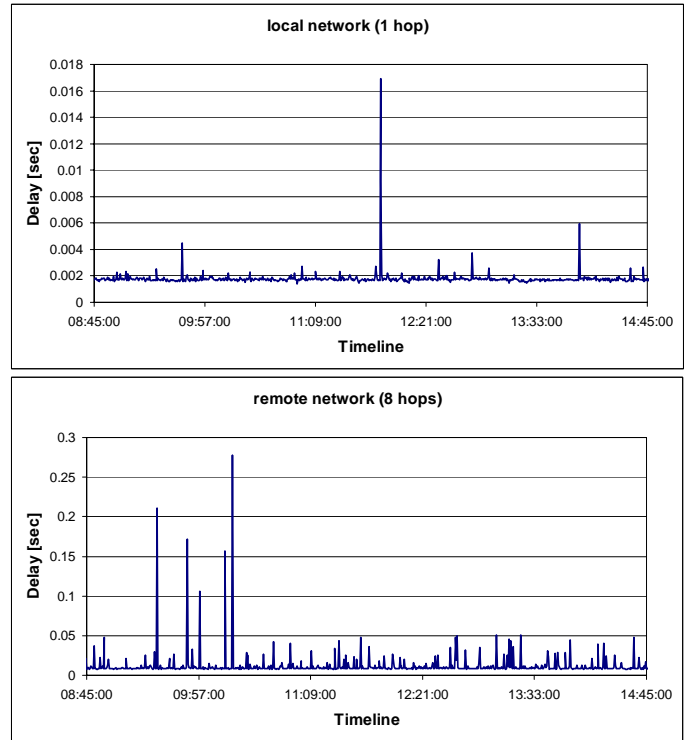

Figure 6: MQM based one-way delay (OWD) measurements within our local network (left) and to a remote network (right)

Similar experiments have been executed to demonstrate the applicability of MQM to indirectly observe the behavior of communication paths. We installed two MQM probes at different locations in Germany, in Erlangen and one in Regensburg and Bayreuth, respectively. Using explicit MQM Ping Requests that were sent from one probe in Erlangen, we were able to estimate the OWD between all four probes. Exemplarily, we show measurement results between Regensburg and Bayreuth and between Regensburg and Erlangen, respectively, in Figure 7. The measured delay values depict the current network quality between the analyzed locations. Less than 20 ms in the first case indicates a high network quality whereas the fluctuating results between 20 ms and 200 ms in the second case represent a measure for network congestion.

Both experiments demonstrate the possibility for delay measurements using the MQM ping mechanism. The inherent
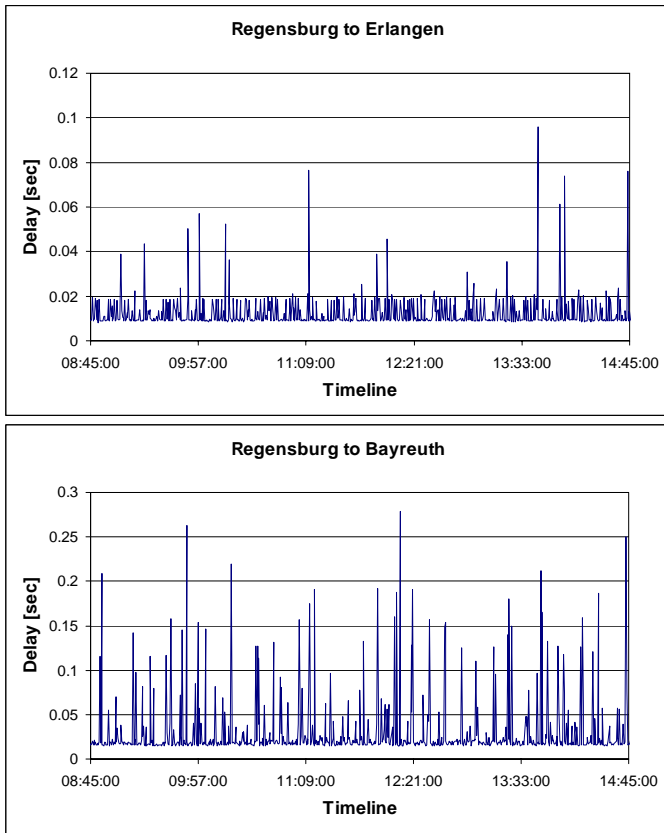
Figure 7: MQM based one-way delay (OWD) measurements from Regensburg to Erlangen (left) and to Bayreuth (right), respectively

capability of multicast to reach multiple destinations with one message was successfully exploited by MQM for these experiments.

5.2 Simulation Results

We used the well-established simulation environment OMNeT++ [18] to implement the MQM model. The INET framework already provided all necessary functionality to create an IP-based network model including multicast routing. Therefore, we only implemented a module MQMApp which directly uses the transport layer module (UDP) provided by OMNeT++. The established network topology is shown in Figure 8. The hosts are running the MQM application while all routers are used to forward IP multicast traffic.

The simulation model allows measurement of several statistics during an experiment. In the following, we discuss the behavior of the MQM ping process, i.e., the number of MQM messages in the network, and present results for the delay measurements. Figure 9 depicts the number of MQM messages in the network. At startup time, each host is required to send a MQM request message. Afterwards, the rule for MQM ping suggests two concurrent messages. Therefore, the graph shows a peak during the first 120 seconds (6 requests). Subsequently, two messages per time interval were recorded. The number of response messages (as shown in the same figure) is correlated to the MQM requests. Each host receiving a request message must send an adequate MQM response. Therefore, 30 MQM responses were recorded during the first 120 seconds. Then, the stationary algorithm shows exactly 10 response messages all the time (2 requests answered by 5 hosts).

To show the capabilities of the MQM (and to show the behavior of the simulated network), some one-way delay measurements are presented in Figure 10. Shown are the results for an experiment that lasted one hour. The transmission delay in the core network (between the routers) was quite small, configured to be exponentially distributed in [0, 10µs]. We configured fast low-latency connections between host1, host2, host3, and host4 to the network. Therefore, the OWD between host1 and host2, host3, and host4, respectively, is quite
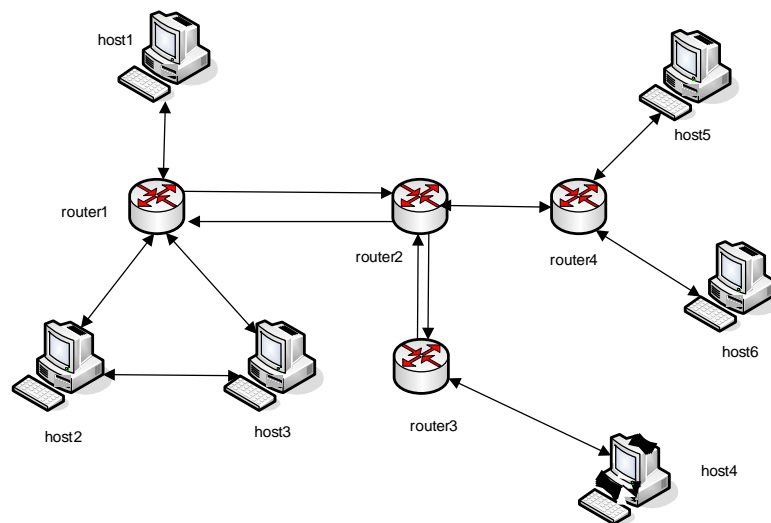

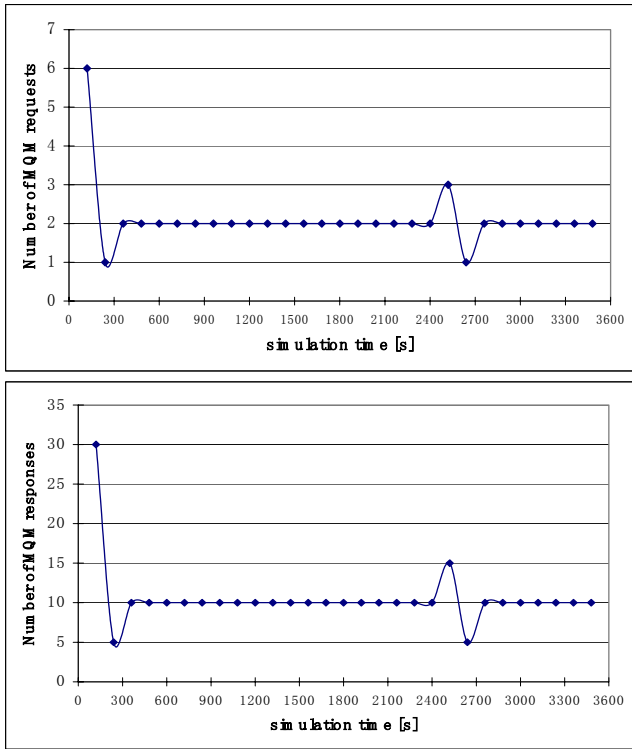Figure 8: Network topology used in the simulation experiments

Figure 9: Number of MQM messages in the network: requests (left) and corresponding responses (right)
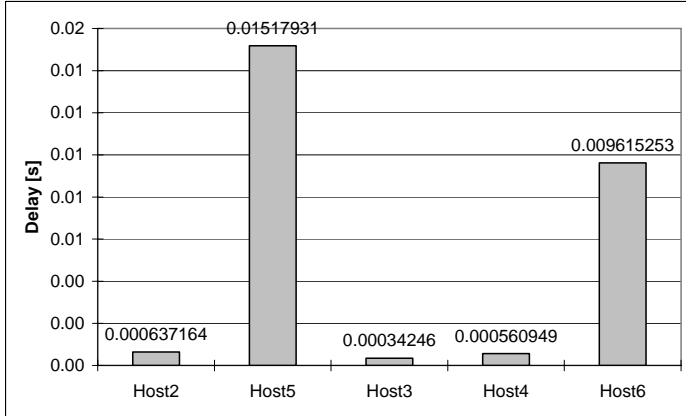

Figure 10: One-way delay measurement in the simulation environment (from host 1)

small (about 0.3 to 0.6ms). In contrast, the latency of the connection between hosts 5 and 6 and the network was configured be distributed in [4, 18ms]. Therefore, the measured one-way delay corresponds with this setup and shows much higher values (9 to 15ms).

6 Conclusion

In conclusion, it can be said that we successfully developed and tested a system for availability analysis of multicast networks that clearly outperforms other solutions. Table 2 shows a comparative overview to the characteristic properties of established multicast measurement tools and the MQM. The basic MQM ping algorithm exploits the multicast properties to duplicate messages within the network. Based on this property, only two active MQM ping requests are sufficient to provide a reliable and complete measurement between all participating MQM probes. The multicast nature ensures that the resulting MQM ping response messages represent the complete network behavior.

The multicast quality monitor includes a new methodology that allows a high scalable reachability analysis which covers all the parts of the possibly split network. The MQM ping mechanism was closely analyzed and was shown that it allows a statement on the availability of single network connections as well as of larger network parts. The high scalability of this approach allows the involvement of a large number of measurement probes without an exorbitant impact on the network.

The final question is "Quo vadis?" The collection of the measurement results can be further optimized. New approaches known from peer-to-peer networks can be employed for this task. Nevertheless, it must be said that the most important problem, the measurement itself, can be assumed to be solved by the new methodology, the MQM ping mechanism.

References

[1] K. Almeroth, L. Wei, and D. Farinacci, "Multicast Reachability Monitor (MRM)," draft-ietf-mboned-mrm-01.txt, July 2000.

[2] K. C. Almeroth, "The Evolution of Multicast: from the MBone to Inter-Domain Multicast to Internet2 Deployment," *IEEE Network Magazine*, 14(1):10-20 January/February 2000.

[3] K. C. Almeroth, K. Sarac, and L. Wei, "Supporting Multicast Management Using the Multicast Reachability

Table 2: Comparison of the capabilities of established multicast measurement tools and MQM

	MRM	Multicast Beacon	MQM
Scalability (smooth adaptation to a large network environment)	Depending on the number of senders	Not at all	Linear with the number of probes
Completeness (capability to detect network partitioning)	Not at all	Detection is possible by the server. This requires a working unicast connection to/from each probe to the server.	Yes, algorithm-inherent property
Flexibility (possibility to measure the current network quality as well)	Limited, only unidirectionally between the probes towards the server.	Yes.	Yes, algorithm-inherent property

Monitor (MRM) Protocol," UCSB, Technical Report TR 2000-26, May 2000.

[4]   K. Chen, M. Kutzko, and T. Rimovsky, "Multicast Beacon Server v0.8.X (Perl)," (http://dast.nlanr.net/projects/beacon), January 2002.

[5]   F. Dressler, "How to Measure Reliability and Quality of IP Multicast Services?," *Proceedings of IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (IEEE PACRIM'01)*, Victoria, B.C., Canada, 2: 401-404, August 2001.

[6]   F. Dressler, "An Approach for QoS Measurements in IP Multicast Networks, MQM - Multicast Quality Monitor," *Proceedings of International Network Conference (INC 2002)*, Plymouth, UK, July 2002.

[7]   F. Dressler, "MQM - Multicast Quality Monitor," *Proceedings of 10th International Conference on Telecommunication Systems, Modeling and Analysis (ICSTM10)*, Monterey, CA, USA, 2:671-678, October 2002.

[8]   F. Dressler, "Availability Analysis in Large Scale Multicast Networks," *Proceedings of 15th IASTED International Conference on Parallel and Distributed Computing and Systems (PDCS 2003)*, Marina del Rey, CA, USA, I:399-403, November 2003.

[9]   N. F. Maxemchuk and D. H. Shur, "An Internet Multicast System for the Stock Market," *ACM Transactions on Computer Systems (TOCS)*, 19(3):384-412, August 2001.

[10] P. V. Mieghem, G. Hooghiemstra, and R. v. d. Hofstad, "On the Efficiency of Multicast," *IEEE/ACM Transactions on Networking (TON)*, 9(6):719-732, Dec. 2001.

[11] P. Namburi, K. Sarac, and K. Almeroth, "Practical Utilities for Monitoring Multicast Service Availability," *Computer Commmunications*, 2005.

[12] T. Oh-ishi, K. Sakai, K. Kikuma, and A. Kurokawa, "Study of the Relationship between Peer-to-Peer Systems and IP Multicasting," *IEEE Communications Magazine*, 41:80-84, January 2003.

[13] M. Ramalho, "Intra- and Inter- Domain Multicast Routing Protocols: A Survey and Taxonomy," *IEEE Communications Surveys & Tutorials*, 13(1):2-25, 2000.

[14] K. Sarac and K. C. Almeroth, "Monitoring Reachability in the Global Multicast Infrastructure," *Proceedings of International Conference on Network Protocols (ICNP)*, Osaka, Japan, 2000.

[15] K. Sarac and K. C. Almeroth, "Supporting the Need for Inter-Domain Multicast Reachability," *Proceedings of Network and Operating Systems Support for Digital Audio and Video (NOSSDAV '00)*, Chapel Hill, NC, USA, June 2000.

[16] K. Sarac and K. Almeroth, "Application Layer Reachability Monitoring for IP Multicast," *Computer Networks*, 48(2):195-213, June 2005.

[17] A. Striegel and G. Manimaran, "A Survey of QoS Multicasting Issues," *IEEE Communications Magazine*, 40(6):82-87, June 2002.

[18] A. Varga, "OMNeT++ Discrete Event Simulation System," *Proceedings of European Simulation Multiconference (ESM'2001)*, Prague, Czech Republic, June 2001.

[19] Tina Wong and Randy Katz, "An Analysis of Multicast Forwarding State Scalability," *Proceedings of the 8th IEEE International Conference on Network Protocols (ICNP 2000)*, Osaka, Japan, pp. 105-115, November 2000.