PGM-based reliable multicast data dissemination in satellite IP networks

Agnieszka Chodorek
Department of Telecommunications and Photonics
Kielce University of Technology
Kielce, Poland
a.chodorek@tu.kielce.pl

Robert R. Chodorek
Department of Telecommunications
The AGH University of Science and Technology
Kraków, Poland
chodorek@kt.agh.edu.pl

***Abstract*— A reliable distribution of information to receivers located in different regions of world (e.g. simultaneous uploading of web servers), can be effectively carried out using multicast technology. From the other side, the usage of satellite IP network (from IP angle of view: a single-hop broadcast link) allows easy disseminate information to a large group of users that may span a large geographical area. In the paper, possibilities of usage of Pragmatic General Multicast (PGM) transport protocol for wide-area distribution using satellite links are discussed. Simulational results show, that PGM is able to assure reliable data transmission and good scalability in heterogeneous terrestrial-satellite environment.**

Keywords-congestion control, performance evaluation, PGM, reliable multicast protocols

I. INTRODUCTION

The usage of satellite IP network allows easy disseminate information to a large group of users that may span a large geographical area. It is especially useful for transmissions that are of a broadcast nature, as a multicast distribution. Although many authors have addressed the problem of real-time multicasting of streaming media via satellite (or mixed satellite-terrestrial) link (see e.g. [1]), the problem of reliable multicast data transmission still remains an unresolved issue.

One of the newest multi-purpose transport protocols intended for reliable multicast transmission is Pragmatic General Multicast (PGM) [2]. PGM is promoted by large producers of software (e.g. Microsoft – PGM is a part of Windows Server 2003 operating system) and network equipment (e.g. Cisco). It provides end-to-end transport service suitable for applications that require ordered or unordered, duplicate-free, multicast data delivery from multiple sources to multiple receivers [2]. The data transport is augmented by PGMCC – a single rate, TCP-friendly multicast congestion control scheme, which uses window-based, TCP-like control loop [3].

The aim of the paper is to analyze the performance of PGM and PGMCC/PGM over heterogeneous medium, including satellite and terrestrial (wired or wireless) links. The paper is organized as follows. In the second section, main problems of reliable multicast transmission in satellite IP network are discussed. In the third section, the brief description of PGM and PGMCC protocols is presented. The fourth section describes simulational experiments, while in the fifth section performance evaluation of PGM (PGMCC/PGM) transport protocol over satellite links is presented. Section six summarizes our experiences.

II. RELIABLE MULTICAST TRANSPORT PROTOCOL OVER SATELLITE LINKS – PERFORMANCE PROBLEMS

In the case of multicast transmission of bulk data over satellite links, the five main performance problems are: latency, asymmetry, transmission errors, congestion and scalability.

A. Latency

The end-to-end latency is a sum of total propagation delay, transmission delay and queuing delay. In the broadband satellite network, the dominant component is the propagation delay [1]. The propagation time between two stations via satellite ranges from tens to hundreds of milliseconds. For instance, the propagation delay calculated for the LEO, MEO and GEO altitudes are of the order of 7, 75 and 260 ms, respectively [4].

In the case of flow controlled transmission, large latency of satellite network can limit throughput achievable by transport protocol. If the multicast transport protocol operates in heterogeneous dissemination system, which integrates satellite and terrestrial networks, performance of the system will be limited be the receiver characterized by the largest end-to-end latency (probably in the satellite branch of the multicast tree).

It’s worth to mention that large end-to-end latency requires large retransmission buffers. Buffers for reliability are sized according the delay-bandwidth product of the network topology.

B. Asymmetry

An important feature of satellite links is bandwidth asymmetry. Satellite networks are asymmetric in at least two ways:

- apparent bandwidth asymmetry – e.g. direct broadcast satellite downlink and return via dial-up modem line,
- unapparent bandwidth asymmetry – e.g. direct broadcast satellite downlink (at Mb/s) and return via slower uplink (at kb/s).

This work was supported by State Committee for Scientific Research (KBN) under Grant No. 4 T11D 015 24 (years 2003-2005).

Bandwidth asymmetry is especially important in the case of reliable transport protocols (multicast or unicast). The closed loop formed by such a protocol between the end systems (feedback of error and/or congestion information through ACK or NACK acknowledgements) results in dependence of achieved throughput on network conditions both on the forward and reverse path.

C. Transmission Errors and Congestion

Packets may be lost due to transmission errors or network congestion. Transmission errors are a function of the characteristics of the satellite, of the earth stations and local environment and interference conditions [5]. In result, technically-identical but geographically-dispersed receivers can have different reception conditions.

Satellite link is characterized by higher bit error rates (BER) than typical terrestrial one. Typical BER on the order of 10^{-7} or 10^{-4} in the worst case [1] is quite enough for analog voice and video services, but it's unacceptable for reliable data transmission. However, due to new modulation and coding techniques, along with highly powered satellites, normal bit errors are usually much lower and achieve "fiber-like" quality [5]. Current satellite systems designed for data transmission generally have BER as low as 10^{-6} or even 10^{-10} [5]. In effect, current systems suffer more likely from congestion than from transmission errors.

RFC 3272 defines congestion as "a state of a network resource in which the traffic incident on the resource exceeds its output capacity over an interval of time" [6]. In the case of satellite networks, congestion is usually build in between the terrestrial and satellite segment of the network (major earth station) while inter-satellite links stays uncongested (or stays at incipient stages of congestion).

D. Scalability

Last but not least problem of reliable multicast transmission over satellite links is multicast session scalability. Reliable transport protocols (multicast or unicast) recovers data that are damaged or lost using retransmission (typically – selective retransmission) of a copy of a damaged (or missing) packet or reconstruct that packet using Forward Error Correction (FEC). The retransmission requires the feedback of packet loss information through returning acknowledgements. In unicast protocols (e.g. TCP), typically positive acknowledgment (ACK) from the receiver are used. If the ACK is not received within a timeout interval, the missing data is retransmitted. In multicast protocols such a signaling method leads to implosion of positive acknowledgments, which are sent from multiple receivers to the source.

Although single ACK has typically smaller size than data packet, the large amount of acknowledgments can cause heavy congestion, especially in the neighborhood of the source or/and in the node at the top of a large branch of multicast delivery tree. In satellite network this effect can be amplified by bandwidth asymmetry. The implosion of ACK is the main limitation of the multicast session scalability – multicast session should be small enough to avoid congestion caused by acknowledgements.

One of possible solution of the problem of implosion of acknowledgements is usage of negative acknowledgements (NACK). The negative acknowledgement is explicit packet loss indication – receiver sends NACK to request retransmission. In the case of lossless and uncongested link, usage of NACKs significantly limits amount of necessary acknowledgements. However, if the packet is lost at the node near the root of a large multicast delivery tree (or in the node at the top of a large branch of the tree), dangerous of a NACK implosion will appear in the network. Thus, reliable multicast transport protocol should apply effective NACK suppression among the receivers and/or NACK aggregation to avoid implosion of negative acknowledgements.

III. AN OVERVIEW OF PGM AND PGMCC PROTOCOLS

Pragmatic General Multicast (PGM) [2][7][8] is a transport protocol designed for reliable distribution of data from multiple sources to multiple receivers. PGM implements reliability by a typical method of selective retransmission, however FEC also is acceptable. The protocol multicastly distributes user data in Original Data (ODATA) packets, while damaged or missing data are retransmitted (multicastly) using special Repair Data (RDATA) packets. To avoid possibility of ACK implosion, PGM uses negative acknowledgments (NAKs) and the dangerous of NAK implosion is reduced by both NAK suppression in receivers and NAK aggregation in PGM routers (so-called PGM NE or PGM-capable Network Elements).

Data are typically retransmitted by the source. However, to constrain retransmission only to certain fragments of distribution tree (and, in result, to improve multicast session scalability) Designated Local Repairers (DLR) can be used. DLR assures local retransmission, so properly located DLR also minimize period of time between transmissions of NAK and RDATA.

Characteristic for PGM is that the transport protocol builds its own distribution tree (PGM tree), which is an overlay network located over the IP routing's multicast distribution tree. If all of routers which build multicast distribution tree are PGM NE, PGM tree will be identical with multicast distribution tree. PGM tree is build using Source Path Messages (SPMs) from a sender, periodically interleaved with ODATA.

PGM may be extended with a congestion control building block – typically with PGMCC [3] (but other building blocks also can be used). PGMCC is a single rate, TCP-friendly multicast congestion control scheme, which uses a window-based, TCP-like control loop. PGMCC emulates TCP's congestion window (cwnd) using token bucket mechanism, which limits amount of data (ODATA, RDATA and SPM packets) a PGM can send. The PGMCC's congestion window is sized using AIMD (Additive Increase Multiplicative Decrease) algorithm and according to feedback from so-called acker. The acker is an "unlucky" user which has the worst reception conditions or – more formally – a receiver, which will achieve the lowest throughput if data transmission utilizes many independent TCP connections. Acker is selected using a TCP throughput equation from [9]:

$$T_i = \frac{1}{RTT_i \cdot \sqrt{p_i} \cdot \left[1 + 9 \cdot p_i \cdot \left(1 + 32 \cdot p_i^{\,2}\right)\right]} \tag{1}$$

where T_i is a TCP throughput for the i^{th} receiver, RTT_i is a round-trip time for the i^{th} receiver and p_i is a packet error rate for the i^{th} receiver.

IV. EXPERIMENTS

Experiments were carried out using Berkeley's ns-2 network simulator [10]. Because build-in PGM model does not support PGMCC, we use alternative model developed by L. Rizzo, available at *http://info.iet.unipi.it/~luigi/pgm.html*. The Rizzo's PGM model was extended using ours DLR model (not implemented in the Rizzo's software, up to now).

We simulate data dissemination system which integrates satellite and terrestrial networks. Analyzed network has double-bottleneck topology (Fig. 1). The first bottleneck is a terrestrial segment 10 Mb/s and the second bottleneck is a satellite link (2 Mb/s). Data must first be transmitted across terrestrial bottleneck and then, if necessary, will be transmitted via satellite. Note, that PGM protocol requires the same forward and return path to achieve full scalability (e.g. DVB-S2 forward path and DVB-RCS return path).

In experiments we explore how PGM packet size, packet error rate (caused by both, transmission errors in satellite link and packet losses due to congestion) and propagation delay impact the protocol's performance. Bulk data (*ftp*) and constant bit rate (CBR 600 kb/s) traffic sources (modeled by build-in models) were used. Data were multicastly distributed from the source S1 to receivers R1…R10. As the source of background traffic, transmitted from S2 to R11, *ftp* over TCP (SACK version) was used. Maximum segment size (MSS) of TCP packet and payload length of PGM packet was set to 1000 B.

Duration of experiment was calculated using typical steady state rule. Simulated transmission time was set according to error rate and was large enough to assure that system is in a steady state (transient components of results are negligible with at least 90% confidence).

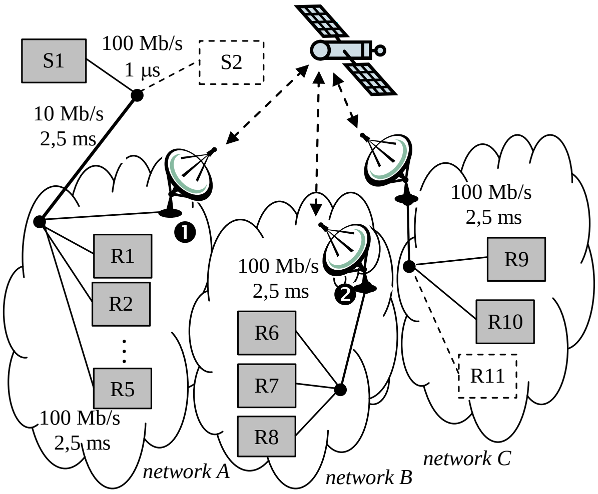

Figure 1. Topology used in experiments

V. PERFORMANCE EVALUATION OF PGM AND PGMCC/PGM OVER SATELLITE NETWORK

In the section, simulation results are presented that illustrate the performance of PGM transmission. Fig. 2a shows goodput of PGM transmission (receiver R6) as a function of PGM payload length (packet size minus header size) in the absence of background traffic. In the case of bulk data transfer (*ftp*) over PGMCC/PGM, large delay of GEO satellite link (about 260 ms) has strong impact on protocol's performance if payload length is too small. If payload length is large (here: larger than 5000 B), the achieved throughput stabilizes on the capacity of bottleneck (satellite) link. In the case of CBR (600 kb/s) transmission, small values of payload length results in large throughputs (e.g., about 900 kb/s for payload length equal to 100). However, it is caused only by large overheads and goodput of transmission stays unchanged (Fig. 2a).

Fig. 2b compares goodput of *ftp* over PGM/PGMCC. The uppermost curve shows the PGM goodput without background traffic, while the two lowermost curves show goodput of PGM competing with TCP. Small rwnd causes that TCP is not able to seriously influence competing PGM flow (the second curve). In result PGM goodput is close to the best-case. If rwnd is large enough and payload length is equal to or smaller than MSS, PGM/PGMCC will show TCP-friendly behavior (the lowermost curve). If payload length exceeds MSS, PGM will become more aggressive and loss TCP-friendliness. It lead to unfair bandwidth allocation – PGM shows tendency to utilize whole available bandwidth.

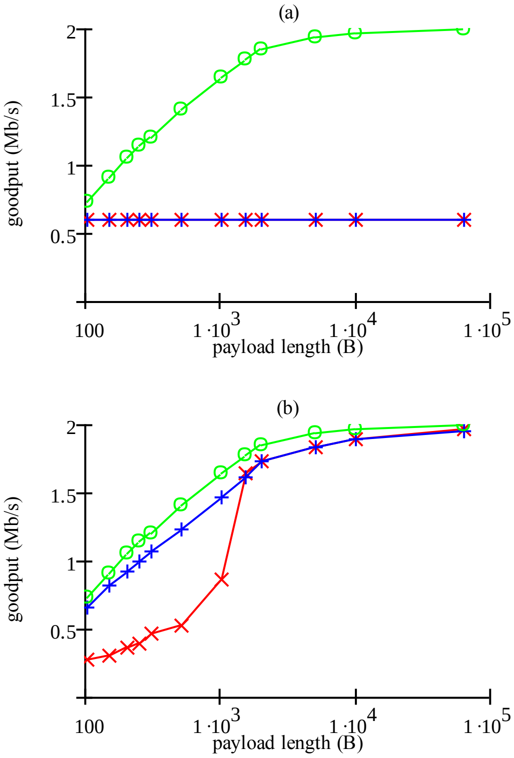

Figure 2. PGM goodput as a function of payload length: a) no background traffic; Legend: CBR over PGM (x), CBR over PGM/PGMCC (+), *ftp* over PGM/PGMCC (o); b) *ftp* over PGM; Legend: TCP traffic, receiver window rwnd = 100 (x), TCP traffic, rwnd = 20 (+), no background traffic (o).

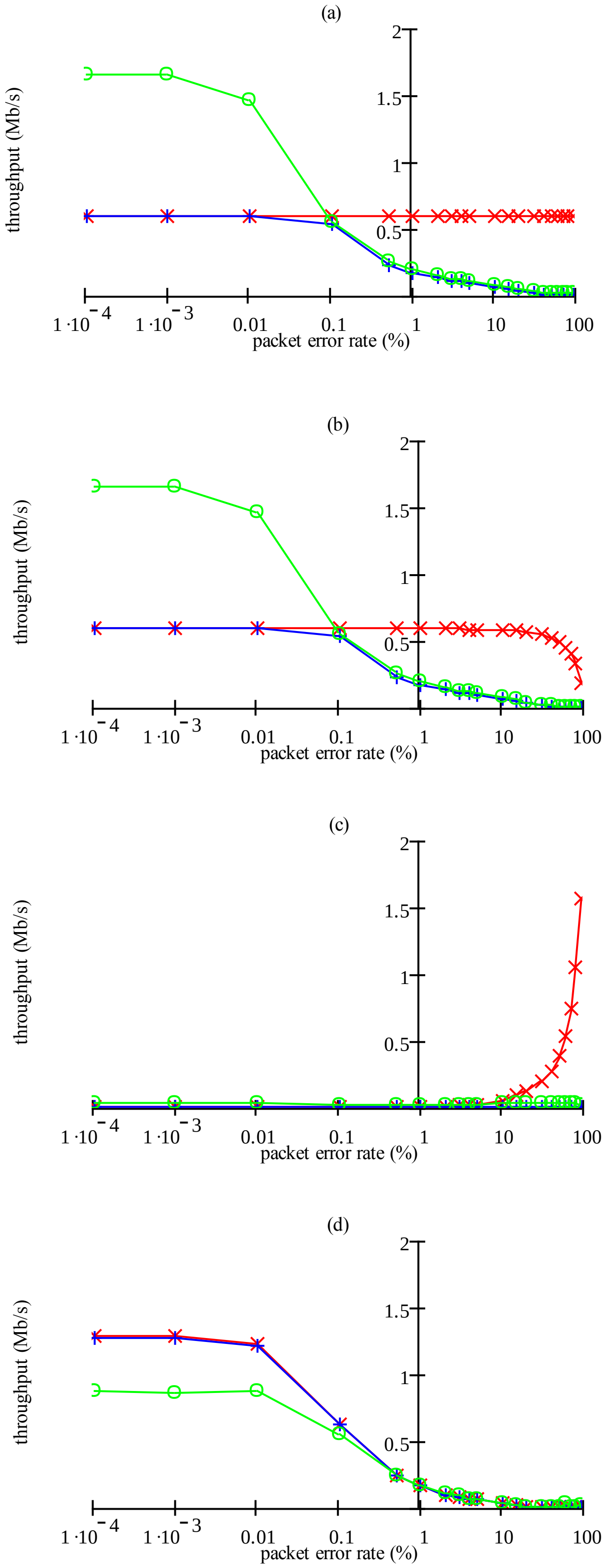

Figure 3. Performance of a system without DLR repairer: a, c) PGM troughput at receiver R1, b) PGM troughput at receiver R6, d) TCP troughput at receiver R11; a, b, d) useful data, c) useless data;
Legend: CBR over PGM (x's), CBR over PGM/PGMCC (+'s), *ftp* over PGM/PGMCC (o's).

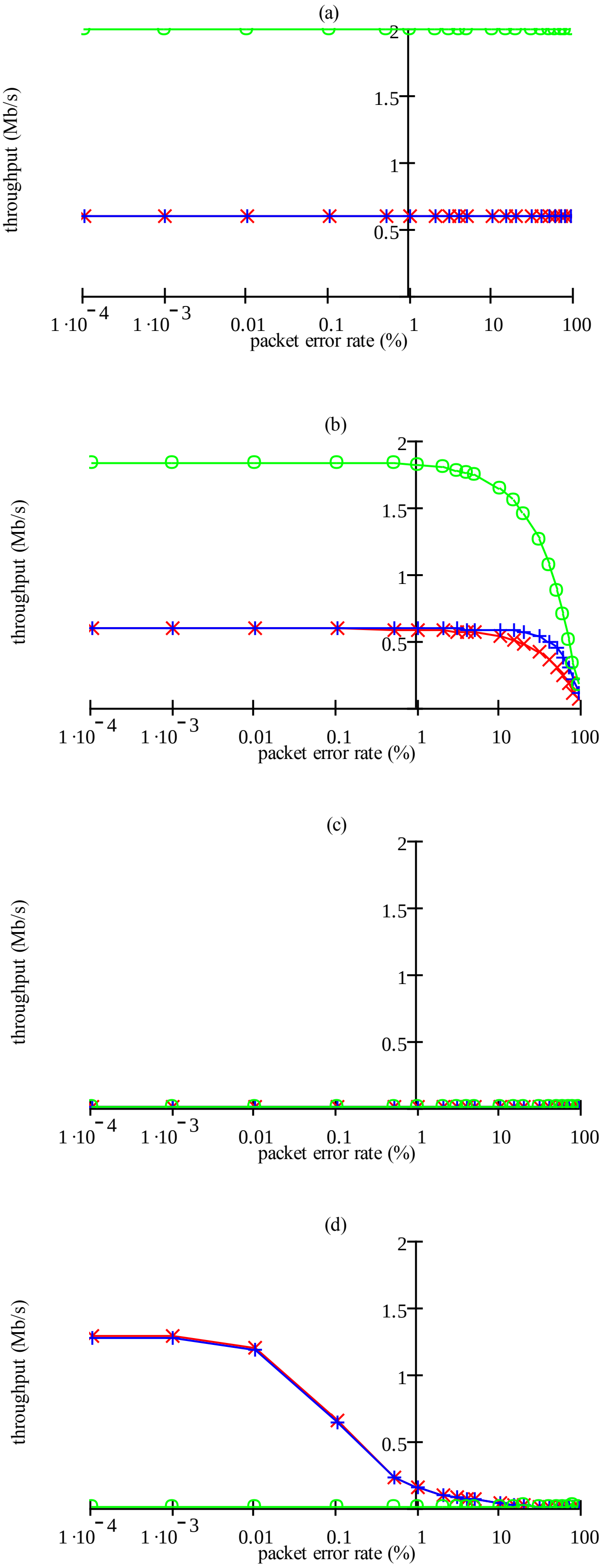

Figure 4. Performance of a system with DLR repairer (location ❶): a, c) PGM troughput at receiver R1, b) PGM troughput at receiver R6, d) TCP troughput at receiver R11; a, b, d) useful data, c) useless data;
Legend: CBR over PGM (x's), CBR over PGM/PGMCC (+'s), *ftp* over PGM/PGMCC (o's).

Fig. 3 and Fig. 4 shows the influence of the packet error rate (PER). Fig. 3 illustrates situation, when there isn’t DLR repairer in a data dissemination system. The first experiment was carried out without background TCP traffic. Fig. 3a illustrates that performance of CBR over PGM doesn’t depend on packet losses which occurs in other fragment of multicast delivery tree. However, the lossy branch of the tree influence the loseless one – RDATA packets appears in the network A. RDATA packets conveys useless (redundant) data what lead to unnecessary high network load. Because of heavy congestion caused by large amount of RDATA (what, in turn, lead to cumulative multiplication of RDATA), if PER exceeds 10%, increase of RDATA throughput is observed in Fig. 3c.

If data transport is augmented by PGMCC congestion control, the problem of cumulative multiplication of RDATA doesn’t exist (Fig. 3c). However, in this situation PGM performance strongly depends on packet losses, regardless of the location of congested node (or lossy link) in delivery tree. It is caused by PGMCC control scheme, in which single acker influence performance of whole data dissemination system. In results, if packet error rate exceeds threshold value, we’ll observe collapse of throughput and doesn’t matter if delivery path from sender to the receiver is uncongested and loseless (Fig. 3a) or is not (Fig. 3b).

The second experiment was carried out with background TCP traffic. If PGM or PGM/PGMCC competes for bandwidth with TCP flow, shape of characteristics of throughput vs. PER do not changes. Values of throughput will be lower (in the case of *ftp* over PGM/PGMCC) or approximately the same (if the network is well-dimensioned for real-time transmission – here: CBR). Fig. 3d depicts TCP throughput as a function of PER. If packet error rate is greater than 0.01%, throughput curve collapses. This collapse is caused by TCP congestion control mechanism and, generally, do not depend on PGM (or PGMCC) behavior.

Fig. 3 shows PGM performance of the network without DLR repairer. DLRs are intended to improve performance of the PGM connections. However, if DLR will be located behind the satellite link (see ❷ in Fig. 1), the repairer is not able to change significantly functionality of the data dissemination system. Thus, characteristics of the system with DLR located at ❷ will be essentially the same as depicted in Fig. 3.

The best results will be observed if DLR repairer is located at the gateway between the terrestrial and satellite segment of the network (location ❶ in Fig. 1). PGM throughput, observed at the receiver R1 in uncongested and loseless network A, depends only on parameters of the source and/or delivery path. PGM throughput in network A doesn’t depend on losses which appear in other fragments of delivery tree, regardless of usage of PGMCC building block (Fig. 4a). Because RDATA are sent by DLR (and not by the source), throughput of RDATA in network A is equal to 0 (Fig. 4c). Receivers located at congested or lossy branches of the delivery tree achieves throughput close to maximal for small and medium PERs (Fig. 4b). In the case of larger packet error rates obtained throughput is much larger than in the system without DLR. However, both PGM and (something surprisingly) PGM/PGMCC transmission manifest strong TCP-infriendly behaviour (Fig. 4d).

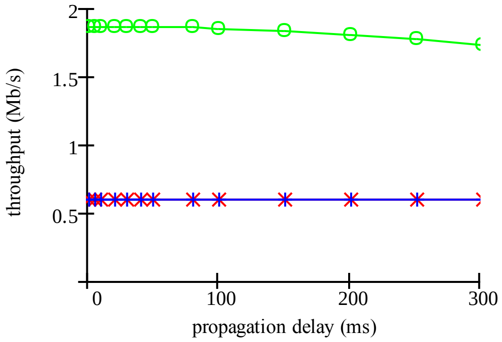

Figure 5. Throughput of PGM transmission received by R6 as a function of propagation delay. Legend: as in Fig. 4.

Fig. 5 shows that performance of both, PGM and PGM/PGMCC, weakly depend on propagation delay. This property is especially important in the case of LEO satellites, where RTT can significantly change during the transmission.

All experiments show good scalability of the multicast transmission with respect to session size and two groups of receivers (group I – network A and group II – networks B and network C).

VI. CONCLUSIONS

PGM is multi-purpose, reliable multicast transport protocol. It can be extended using PGMCC building block, which provides TCP-like congestion control. PGMCC limits transfer rate according to the worst-case receiver. Experiments, carried out in Berkeley’s ns-2, show that PGM is able to assure reliable data transmission and good scalability in heterogeneous terrestrial-satellite environment.

REFERENCES

[1] Internetworking and Computing over Satellite Networks, Y. Zhang, Ed. Boston: Kluwer Academic Publishers, 2003.

[2] T. Speakman, J. Crowcroft, J. Gemmell, D. Farinacci, S. Lin, D. Leshchiner, M. Luby, T. Montgomery, L. Rizzo, A. Tweedly, N. Bhaskar, R. Edmonstone, R. Sumanasekera, and L. Vicisano, “PGM Reliable Transport Protocol Specification,” RFC 3208, December 2001.

[3] L. Rizzo, G. Iannaccone, L. Vicisano and M. Handley, “PGMCC single rate multicast congestion control: Protocol Specification,” INTERNET-DRAFT, Work in progress, July 2004.

[4] B. R. Elbert, The Satellite Communication Ground Segment and Earth Station Handbook. Boston, Artech House, 2001.

[5] Handbook on Satellite Communications. 3rd ed., New York, Wiley, 2002.

[6] D. Awduche, A. Chiu, A. Elwalid, I. Widjaja, and X. Xiao, “Overview and Principles of Internet Traffic Engineering,” RFC 3272, May 2002.

[7] J. Gemmell, T. Montgomery, T. Speakman, N. Bhaskar, and J. Crowcroft, “The PGM Reliable Multicast Protocol”,. IEEE Network, vol. 17, no. 1, January 2003.

[8] “The Reliable Multicast Protocol Component of Windows Server 2003,” The Cable Guy, June 2003.

[9] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose, “Modeling TCP Throughput: A Simple Model and its Empirical Validation,” Proc. of ACM SIGCOMM 1998.

[10] The ns-2 simulator. Available: http://www.isi.edu/nsnam/ns/.