

A Scalability Scheme for the Real-time Control Protocol

R. El-Marakby, D. Hutchison

Computing Dept., Lancaster University, Lancaster LA1 4YR, U.K.

Tel.: (+44)-1524-65201 Ext. 94538, Fax: (+44)-1524-593608

E-mail: <randa,dh>@comp.lancs.ac.uk

Abstract

Recently, some problems related to using the Real-time Control Protocol (RTCP) in very large dynamic groups have arisen. Some of these problems are: feedback delay, increasing storage state at every member, and ineffective RTCP bandwidth usage, especially for receivers that obtain incoming RTCP reports through low bandwidth links. In addition, the functionality of some fields (e.g. packet loss fraction) in the Receiver Reports (RRs) becomes questionable as, currently, an increasing number of real-time adaptive applications are using receiver-based rate adaptive schemes instead of rate adaptation schemes based on the sender.

This paper presents the design of a scalable RTCP (S-RTCP) scheme. S-RTCP is based on a hierarchical structure in which members are grouped into local regions. For every region, there is an Aggregator (AG) which receives the RRs sent by its local members. The AG extracts and summarises important information in the RRs, derives some statistics, and sends them to a Manager. The Manager performs additional statistical analysis to monitor the transmission quality and to estimate regions which are suffering massively from congestion.

We believe that our S-RTCP alleviates some of the RTCP scalability problems encountered in very large dynamic groups and makes effective use of RRs with regard to the current changing requirements of real-time adaptive applications in the Internet today.

Keywords

RTCP scalability, RR, TTL, AG, LAG, AGR, Manager

The original version of this chapter was revised: The copyright line was incorrect. This has been corrected. The Erratum to this chapter is available at DOI: [10.1007/978-0-387-35388-3_42](https://doi.org/10.1007/978-0-387-35388-3_42)

1 INTRODUCTION

Today, the Real-time Transport Protocol (RTP) is widely deployed in most Mbone applications over the Internet involving multiple senders and receivers. The Real-time Control Protocol (RTCP) which is RTP's control protocol is used mainly in adaptive applications where the sender changes its rate of data transmission in order to suit the current state of the network.

Some problems have arisen when RTCP has been used in very large dynamic multicast groups (Rosenberg, 1997a), (Rosenberg, 1997b), (Schulzrinne, 1997). Firstly, because the RTCP reporting interval grows linearly with the group size, a feedback delay occurs. Consequently, infrequent feedback reports are sent and timely reporting does not occur. Second, each member has to keep track of every other member in the group, thus a storage scalability problem can appear. Third, a flood of initial RTCP reports multicast to the whole group can occur when large number of members join at the same time. As a result, members can be flooded with these reports, especially the ones connected to the network through low bandwidth links, and the network may be congested. This problem occurs (Aboba, 1996) if the reporting members are not implementing the reconsideration algorithm described in (Rosenberg, 1997a). Fourth, for receivers connected through low bandwidth links, the RTCP bandwidth available could be used more effectively than is presently the case.

Today, in the Internet, some of the requirements for real-time adaptive applications are changing. An increasing number of the current multicast applications prefer to use receiver-based rate adaptation schemes instead of sender-based rate adaptation schemes to adapt to congestion in the network. In sender-based rate adaptations, when congestion occurs, the sender decreases its rate of data transmission to suit the receiver with the lowest capabilities. Receiver-based adaptive applications have the advantage of accommodating to the heterogeneous capabilities and conflicting bandwidth requirements of different receivers in the same multicast group (McCanne, 1996). Also, adaptation is done immediately instead of waiting for the sender to adapt. With the appearance of these kinds of receiver-based applications, we ask the question: what is the function of the RTCP Receiver Reports (RRs) and how can RRs be used effectively in the current Internet?

We designed a hierarchical scheme which groups members in local regions. Members in each local region send their RRs locally to an Aggregator (AG) in the same region. The AG summarises important information in the RRs, derives some statistics, then sends this information to a Manager. The Manager does some monitoring and diagnosis functions to estimate which regions are suffering highly from congestion and to evaluate the quality of the transmitted data.

This paper proceeds as follows. In Section 2, some background information about RTP/RTCP functionality is presented. Section 3 presents some of the scalability problems of RTCP feedback reports. In Section 4, we describe our Scalable RTCP (S-RTCP) scheme. Section 5 presents some of the benefits of using S-RTCP. Finally, we summarise the current status and outline future work. The present paper expands on the rationale and description given in (El-Marakby, 1998).

2 BACKGROUND

RTP, the Real-time Transport Protocol, is mainly used for real-time transmission of audio and video over the Internet in multicast and unicast modes (Schulzrinne, 1996). It provides several functions:

- Identification of payload data type to identify the format of the payload data.
- Sequence numbering to detect data packet loss and out-of-order packets.
- Timestamping so that data is played out at the right speeds (Rosenberg, 1997a).

RTCP, RTP's Real Time Control Protocol, is used in monitoring the Quality of Service (QoS) of data delivery and in conveying minimal session control information to all members in an audio/video RTP session.

Both RTP and RTCP are integrated within the application such as the Mbone video and audio applications (e.g. vic and vat).

RTCP has five types of report that are periodically sent to all members in the session. The most important are the feedback reports, namely the Sender Report (SR), and the Receiver Report (RR). SR and RR differ only in that the SR is issued by a receiver which is also a sender whereas the RR is issued by a receiver which is not a sender. Both SR and RR contain performance statistics on the total number of packets lost since the beginning of transmission, the fraction of packet loss in the interval between sending this feedback report and sending the previous one, the highest sequence number received, jitter, and other delay measurements to calculate the round-trip feedback delay time. The SR provides more statistics summarising data transmission from the sender, e.g. timestamps, count of RTP data packets, and number of payload octets transmitted.

The SR and RR have several functions. RRs are used mainly in sender-based adaptive applications. The sender can modify its transmissions dynamically based on the RR feedback it receives from its receivers. The packet loss parameter in the RRs has been used as an indicator of congestion in the network. So, after receiving the RRs from the receivers, the sender may increase or decrease its rate of data transmission according to the packet loss fraction it received within the current interval. This rate adaptation helps to reduce network congestion and adapts to changing network conditions (Bolot, 1994), (Busse, 1996), (El-Marakby, 1997a), (El-Marakby, 1997b). The SR is useful in lip-synchronisation (inter-media synchronisation) and in calculating transmission bit rates. Both SR and RR feedback can be used by a third-party monitor which does not receive RTP data but only RTCP packets. This monitor can be an Internet Service Provider (ISP) or a network administrator. It monitors performance of the network and diagnoses its problems (Schulzrinne, 1996).

The other three types are the Bye report which is used when a member is leaving the session, the Application-defined RTCP packet (APP) report which is used for experimental use with no official packet type registration, and the Source Description (SDS) report which provides identification information about all members in the session.

In the next section, we shall describe some of the RTCP scalability problems. Then we will discuss the functionality of RRs with respect to current requirements

of real-time adaptive applications in the Internet today.

3 RTCP SCALABILITY PROBLEMS AND FUNCTION OF RRs

The feedback provided by Internet applications has proved to be useful as no special support is needed from the network to detect its current state. The RTCP feedback is used in adaptive applications as well as in monitoring.

RTCP scales well for small multicast groups but a scalability problem arises when it comes to a group of *thousands* of users. Some of these problems are addressed in (Rosenberg, 1997a), (Rosenberg, 1997b).

3.1 RTCP feedback problems in a large dynamic group

We will explain first how the interval between RTCP packet transmissions is calculated. All RTCP reports multicast to all members in the group must not consume more than a small fraction (nominally 5%) of the whole bandwidth assigned for the session (Schulzrinne, 1996). Hence, every member has to store an estimate of the size of the group by counting distinct RTCP reports sent to the multicast group. Consequently, members scale back their RTCP reporting interval based on the group size they calculated. That is to say, as the group size increases, each member increases its reporting interval and as the group size decreases, every member decreases its reporting interval. As a result, the bandwidth limit for RTCP reports does not exceed 5% of the whole session bandwidth regardless of changes in the group size at any time during the session.

The following are some of the problems encountered when using RTCP feedback in a large dynamic group:

Feedback delay

The feedback should be sent periodically within acceptable time intervals. In a large RTCP group, this does not happen. Feedback is sent very rarely or not at all. This happens because the RTCP reporting interval grows linearly with the group size. So, as the group size increases, the RTCP interval increases resulting in infrequent RTCP feedback reports which decreases the significance and value of the feedback (Rosenberg, 1997a).

Increasing storage state

In order to calculate the size of the group, every member has to store a count of distinct members it heard from during the session (Rosenberg, 1997a). So as not to count duplicate members, the unique Synchronisation Source identifier (SSRC) found in the RTP header is stored for every distinct member. Of course, storing all the distinct SSRC identifiers for a large group causes a storage scalability problem for every member.

This problem was discussed in (Schulzrinne, 1997) where a SSRC sampling algorithm is described.

Multicasting RRs to the whole group

RRs are multicast to every member in the session. As mentioned before, RRs are mainly used by the senders for adaptation. So, it seems there is very little benefit having each member send its RRs to every other member in the group which are not senders. In addition, members connected to low bandwidth links would not want part of their bandwidth to be used by incoming RRs when this bandwidth usage would be of little or no advantage to them. Moreover, the processing load at every member may increase because of processing incoming RRs from other receivers. Furthermore, if congestion occurs in the network, it is more likely to affect local members near the congested link. Hence, their RR feedback reports will be more or less similar and hence by decreasing the number of redundant RRs that are multicast, congestion can be reduced (Aboba, 1996).

Initial feedback flood

When a very large number of members join simultaneously (e.g. at the start of a MBone multicast session announcement) (Rosenberg, 1997a), it will not be possible to get an accurate estimate of the group size. Each member's first estimation of the group size is 1, and so all the RTCP reports are sent within a fixed initial interval. Consequently, congestion can occur in the network and especially at low bandwidth links of some members. In addition, the feedback reports sent by other members may be dropped due to congestion. This results in inaccurate estimation of the group size which depends on counting the reports coming from distinct members. Hence, it will take a long time to converge to a fairly accurate estimation of the group size and thus to an appropriate RTCP interval computation.

This problem of *initial* RTCP feedback reports was solved by Rosenberg and Schulzrinne (Rosenberg, 1997a), by applying a reconsideration algorithm. Members listen to other members in the group before sending their *initial* feedback reports. Consequently, the reporting interval is readjusted before sending the first feedback report.

Bye flood

When a RTP member leaves the group, it multicasts a RTCP bye packet to the whole group. The problem occurs if many users leave the group at the same time. As a result, a flood of Bye packets that may congest the network occurs. The problem was fixed in (Rosenberg, 1997b) by applying a Bye reconsideration algorithm.

3.2 Functionality of RRs

The Internet is a heterogeneous network. Network resources are varied throughout, and users can have different capabilities specifically link bandwidth. One of the most important functions of RR feedback reports is their usage in adaptive applications. By using the *packet loss fraction* in RRs, the sender can detect network congestion. Hence, the sender changes its rate of data transmission to adapt to changing network conditions and to help reduce congestion. This technique has proved to be useful for unicast applications. However, for multicast applications in the heterogeneous

Internet, the sender ends up decreasing its data transmission rate to suit the receiver with the lowest capability. Consequently, the sender will not be able to meet the various bandwidth requirements of different receivers in the same multicast group.

To accommodate to this heterogeneous environment and to scale to very large number of receivers, a receiver-based rate adaptation scheme is used (McCanne, 1996). The sender sends the data on separate multicast groups. The groups are ordered such that each provides refinement information over the previous group to give increased quality. The receivers can subscribe to one up to all of these multicast groups according to each one's capabilities and according to the current state of the network. Hence, receivers adapt to congestion by joining or leaving multicast groups.

Nowadays, in the Internet, we see a great movement towards receiver-based adaptation schemes. So, in these applications, the packet loss parameter, which is the most significant parameter in RRs, becomes of less or no significance to the sender as adaptations are performed from the receivers immediately without waiting for the sender to react.

In the next section, we describe the architecture of our scheme.

4 OUR SCALABLE RTCP SCHEME

In this section, we describe how locally scoped regions are formed in a hierarchical way. Then, we explain in details the functionality of the Manager.

4.1 Overall view

RTCP feedback reports are multicast mainly for receivers to calculate the group size and thus compute their RTCP reporting interval. In our scheme, the members do not need to compute the whole size of the multicast group and RRs are not multicast.

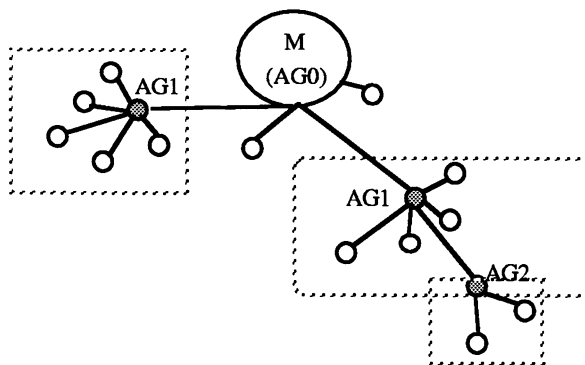


Figure 1 Structure of our scheme showing members in local regions with an AG (shaded circle) per region and a Manager (M) at the root of the hierarchy.

Figure 1 depicts the structure of our scheme which organises members

dynamically in a multi-level hierarchy of local regions. Each region has an aggregator (AG). Local members send their RR feedback reports with limited scope to reach their own AG which gathers and aggregates statistics from these reports which it passes to a Manager. The Manager computes additional statistics to evaluate the transmission quality and to estimate the regions which suffer from congestion.

Our scheme makes use of the Time-to-Live (TTL)* field in the IP header to allow us to build the multi-level hierarchy with locally scoped regions. We are aware of the problems when using TTL scoping with the Distance Vector Multicast Routing Protocol (DVMRP) (Meyer, 1997). We chose to use TTL scoping because it is simple.

4.2 Scheme entities:

The following are the entities of our Scalable RTCP (S-RTCP) scheme:

- Member: A member is a sender or a receiver in the same RTP session.
- LAN Aggregator (LAG): The aggregator for a LAN is also a member which represents *only* local members in the LAN. It aggregates RRs from members in its LAN; it then reports to the Manager.
- Aggregator (AG): The AG is also a member, but it also aggregates RRs from members in its local region (i.e. its children); it then reports to the Manager. The children of an AG can be normal members, AGs, or LAGs. Every AG has a level in the hierarchy. For example, in Figure 1, AG1 is an AG of level 1, while AG2 is an AG of level 2 which is a child of AG1.
- Aggregator Report (AGR): This is a new RTCP report of type AGR. Every AG/LAG sends AGRs to the Manager to summarise the quality received by local members during different intervals.
- Manager: This performs some monitoring and diagnosis functions. It receives AGRs. It is also an AG of level 0 (AG0) as it is at the root of the hierarchy. It receives RR feedback from its direct children who are neither AGs nor LAGs, while it receives AGRs from all the other aggregators in the hierarchy. The Manager should be connected to the network through a fast bandwidth link.

The following subsection provides a detailed explanation of the mechanisms of our scheme.

4.3 Scheme description

When starting the RTP session, two multicast addresses are announced; the first address is for the delivery of RTP data packets, while the second one is for transporting RTCP control packets. Then, the Manager joins the control multicast group. It receives only the RTCP control packets and not the data packets. It is also the first AG in the multi-level hierarchy (AG0). Afterwards, senders and receivers

* This is an integer field in the IP packet header for constraining the travelling distance of the packet. The source initialises the TTL field with an appropriate initial value according to the distance it wants the packet to travel. Each router decrements this TTL by 1 when the packet arrives. The router discards the packet if the TTL reaches zero.

join the two multicast groups for data and control.

Figure 2 depicts the multi-level hierarchical structure of local regions.

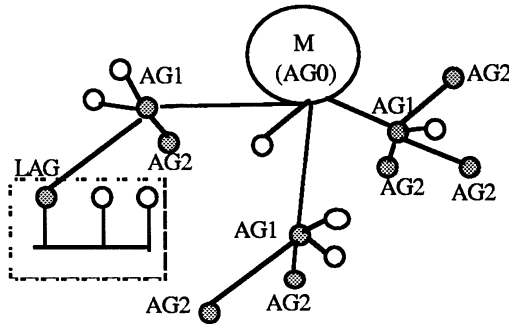


Figure 2 Organisation of members in local regions where each local region has an AG of a certain level in the multi-level hierarchy. The shaded circles represent AGs, while blank ones represent normal children. Also, a LAG (connected to a LAN) is shown.

Selection of Parent AG and formation of local regions

A new member will perform an expanded ring search (Yavatkar, 1995). The new member will repeatedly search for a Parent AG by increasing the TTL value until it finds a near Parent. First, it multicasts a "Search_for_Parent" request (see figure 3) with a small TTL value greater than 1 as it is a well known convention that a multicast packet with TTL=1 is sent only to members in the same LAN. If no reply was received after some time, then it will do another expanded ring search but with a greater TTL value, and so on, until it receives reply(ies) from existing AG(s) of which one will be its future Parent AG.

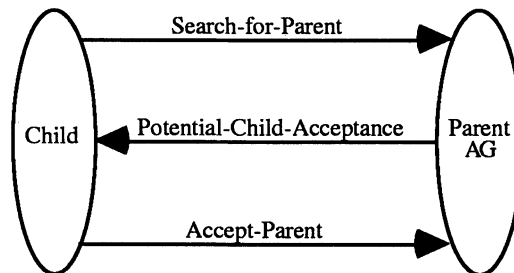


Figure 3 Messages interchanged between child and Parent during the process of searching for a Parent.

Each Parent AG stores the current number of its direct children which includes children acting as AGs or normal children (i.e. not AGs). In addition, each Parent stores the maximum number of children, it is allowed to have, initially obtained from the Manager. That is to say, when the very first new members join at the beginning of the session, the only Parent AG by that time is the Manager which is

AG0. If these new members become AGs, the Manager will pass to them these numbers and other information. Then, these AGs might become parents and pass the information they obtained initially from the Manager to their children AGs and so on.

Upon the receipt of the request "Search_for_Parent", every *Candidate Parent AG* runs the following tests:

Test 1:

If this *Candidate Parent* can afford to get more children (i.e. it currently has fewer children than the maximum number stored)

then

This new member is considered as a *Potential Child*;
Go to Test 2.

else

It will not send any reply back to the new member;
Exit (i.e. Test 2 is skipped).

Test 2:

If (the level of this *Candidate Parent* < N)⁺ AND
(the distance from this new member > certain threshold)

then

This member is considered as a *Potential Child AG*.

else

Exit.

The following is a detailed explanation:

- In Test 2, the *Candidate Parent* checks whether its level in the hierarchy is less than N. If its level is N, this means that the new member cannot be an AG. Consequently, the height of the tree will be limited. We estimate that a suitable height is N=3. In addition, the *Candidate Parent* checks if the distance between the new member and itself is greater than a certain value; if greater, then this new member can be an AG, otherwise it will remain to be just a child. This value should not be small as we do not want the new AG to be very close to its Parent AG. This value is passed initially by the Manager to AGs of level 1 which pass it to AGs of level 2 and so on.
- Note also, if an AG belongs to that Parent but does not have any children, then the Parent can replace a new AG instead of that old one which returns to being a normal child. This is not mentioned in Test 2 for simplicity.
- If a LAG wants to join, it will be accepted right away by the Parent AG no matter how many children it has.

After performing these tests at the Parent, if this new member is a *Potential Child*, then the Parent will send a reply "Potential_Child_Acceptance". The reply will contain the Parent's IP address and a TTL value of 255. Furthermore, if this new member is a *Potential Child AG*, then the reply will include also the Parent's level in the hierarchy, the maximum number of children this new *Potential AG* will be able to have, the minimum distance allowed between this new member and its future direct AGs, and the minimum and the maximum thresholds for measuring

⁺ N is the height of the hierarchical tree. It is equal to the level of the tree + 1.

packet loss (to be explained later).

If there is more than one *Candidate Parent* for this new member, i.e. it receives more than one reply, then it will choose its Parent as the one whose reply carries the largest TTL (i.e. shortest distance from Parent to this new member). That is, as mentioned before, every *Candidate Parent* sends an initial TTL value of 255 in its "Potential_Child_Acceptance" reply so that the child can choose that Parent whose reply carries the remaining largest TTL. Note that in the current phase of our work we are measuring the distance in terms of number of hops only and not delay time. The new member will store all the values in the "Potential_Child_Acceptance" reply sent by the selected Parent. If more than one reply is received from different Parents, the new member will send "Reject_Parent" to those Parents not selected.

Then, the new member will unicast "Accept_Parent" reply to the selected Parent. The Parent will store the remaining value from the original TTL sent by the child in "Accept_Parent" as an indicator of its distance from that child. Hence, it can detect the distance to the furthest child. In addition, it will increase the number of its children by one.

This restriction of the maximum number of children is an attempt to balance the load of the members among local regions. The Manager is the only AG that does not have this restriction. So if all AGs have their maximum number of children, then any new member will be a child of the Manager. In addition, LAGs do not have restrictions on the number of members in their LANs.

Hence, most members may end up in the vicinity of their nearest AG but this is not always the case.

AG leaving or crashed

Every Parent AG multicasts periodically a local refreshment message to its children with TTL=the stored TTL of the furthest child from it. This message shows that the Parent is still alive and not crashed. If the Parent wants to leave the group, it will multicast locally a Bye packet. Whether the Parent crashed or is leaving, every child will start again the process of searching for a Parent AG through expanded ring searching.

An exceptional case arises when the child is an AG. As mentioned before, every AG can have a maximum number of children. In addition, every AG can accept a maximum number of additional *children AGs* that are not its own children only if their Parent crashes or leaves. Hence, if a Parent crashes or leaves the group, then every child AG of this Parent will search for the nearest Parent. If the nearest Parent can accept more AGs of other Parents, then this child AG will take it as its new Parent, otherwise the child AG will expand its ring searching scope to search for another Parent. Note that this exceptional test was not mentioned before for simplicity.

Choice of a LAN Aggregator (LAG)

If one or more members in the same LAN are participating in the same RTP session, a LAN Aggregator (LAG) is chosen to aggregate information from all the members in the LAN. The process of choosing a LAG depends on scoped multicast

discovery queries to locate a LAG for the LAN.

When a new member in the LAN joins the session, it will send a query packet "Search_for_LAG" to search for an existing aggregator for this LAN. If this exists, the LAG will send a reply "LAG_Exists" that contains its IP address to be stored afterwards by the new member.

If no reply is received within some time, then the new member will consider itself to be the first member in this LAN for this RTP session and will elect itself as the LAG. Then, it starts searching for a Parent AG (see the previous subsections). Afterwards, the Parent AG will pass to the new LAG the minimum and the maximum thresholds for measuring packet loss. These parameters are used when summarising RRs received from members in its LAN.

LAG crashing or leaving

The LAG multicasts periodically "LAG_Exists" refreshment message to other members in the LAN to inform them that it is alive (Papadopoulos, 1998). If this message is not received within some time, the LAN members will assume that their LAG crashed.

If the LAG leaves the group, it will multicast locally a RTCP Bye message. It will also unicast to its Parent AG that it is leaving.

Whether the LAG crashed or left the group, each member will start the process of choosing a new LAG for their LAN. Each member will try to multicast locally a "Want_to_be_LAG" request. Each member will use a randomised back-off timer and when the timer expires for one of the members, it will immediately multicast locally a "LAG_Exists" message containing its IP address. Upon receipt of this message, the other members in the LAN will suppress their "Want_to_be_LAG" request and accept this member as their new LAG, then store its IP address. This randomised back-off scheme prevents the flood of the "Want_to_be_LAG" requests if all members multicast at the same time and resolves the problem of choosing a new LAG by directly selecting the LAG whose timer expires first.

SR and Bye RTCP reports

Our scheme deals mainly with the Receiver Reports (RRs). By limiting their travelling distance and summarising important statistics they include, we improve the RTCP scalability. The Sender Reports (SRs) will still be multicast periodically from the sender to the receivers in the session. Note that SRs will not include receiver reporting within them.

Bye reports are sent as follows:

- If a child is leaving, it will send a Bye packet to the Parent AG.
- If an AG/LAG is leaving, it will multicast locally to its children a Bye packet.

4.4 Contents of an AGR

Each AG receives the RR feedback from its direct children which are not AGs or which are AGs with no children. RRs are sent by local members within a certain time interval that is randomised but not to exceed some fixed amount of time. The

AG organises the information, derives some statistics, and includes them into an AGR which reports the quality received by the receivers within a certain interval. Note that the statistics are computed from the RRs of every receiver to a specific sender in the multicast data group.

The QoS statistics and other information contained in an AGR are described below. Some of the functionality of these statistics is explained in the following subsection. The statistics are:

- Number of children that this Parent AG is summarising in the current interval. This number includes only children that send RR feedback.
- Number of children, within the current interval and from the beginning of the data transmission, whose:
 - packet loss exceeds the maximum threshold;
 - packet loss lies between the minimum and the maximum thresholds.
- IP address of the Parent of this AG.
- IP address of the child receiving the worst quality (i.e. which has the highest packet loss in this interval) and the value of packet loss incurred.
- Average, median, and standard deviation of packet loss, in the current interval and since starting transmission, that:
 - surpasses the maximum threshold;
 - lies between the minimum and the maximum thresholds. Note that in order to calculate the median, the AG will sort the packet loss according to the maximum packet loss incurred by every child.

Once all these measurements are computed, they are included in an AGR. Note that these measurements are to evaluate the quality received from one sender. If another sender exists during this interval, then the same measurements are calculated from RRs of local receivers of this other sender. Then, the measurements are appended to the AGR but related to the other sender.

Then, the AG unicasts the AGR to the Manager. In case the aggregator is a LAG, it will send its AGR directly to the Manager too. In the current phase of our work, the AGR is sent directly to the Manager and not to the Parent AG that can pass it to its Parent AG and so on until it reaches the Manager. This is because we do not want to have a long feedback delay until the AGR reaches the Manager.

The following subsection describes the functionality of the Manager.

4.5 Functionality of the Manager

The Manager monitors the data distribution in the multicast group and performs some diagnosis functions. It collects and parses the information received from the AGRs during every interval. Then, it logs useful statistics. By making use of information in AGRs, it can estimate whether problems are specific to a certain region or several regions or to all regions of the whole multicast group.

By obtaining the number of children whose packet loss exceeds the maximum threshold as well as the total number of children in the region, the percentage of children suffering from maximum packet loss is derived. As a result, the Manager can pinpoint the regions which are suffering severely from high packet loss. This percentage can turn out to be the same as in another region. However, the case of

one member suffering from maximum packet loss out of a total of 5 members in the region is different than the case of 100 members suffering from maximum packet loss out of 500 in the other region.

Moreover, the mean, the median, and the standard deviation of packet loss that lies between minimum and maximum threshold and that is incurred by all local members, can be used by the Manager. The Manager computes the distribution of packet loss and hence can detect whether packet loss from most members in this range lies nearer to the minimum threshold or nearer to the maximum threshold or in between. The same derivations apply to packet loss greater than maximum threshold.

Every AGR contains the IP address of the Parent of the AG which is sending this AGR. Consequently, the Manager can trace back congestion and detect if it is also spreading in other neighbouring regions (i.e. region of the Parent) or if it is only limited to the current region.

The IP address of the receiver that is suffering from the maximum packet loss might be used by the Manager to launch an *mtrace* between the sender and the receiver to diagnose network problems along the multicast distribution tree and to detect hops showing a significant amount of losses (Thaler, 1997).

In some cases, if some applications still insist on using sender-based adaptive schemes, then S-RTCP can be adapted so that the Manager sends the packet loss value of the receiver suffering most from the highest packet loss incurred in the current interval to the sender. The sender may decrease its rate of data transmission if necessary.

By storing statistics about several consecutive intervals, it can be detected whether the network performance is improving or not.

The estimations mentioned above are derived from short-term statistics (i.e. statistics within an interval). Moreover, similar analysis can be performed on long term statistics to evaluate the quality of the distribution of data during the whole session.

In addition, the statistical data which is gathered and analysed can be used by an Internet Service Provider (ISP), a network administrator, or a technician to estimate the quality received by each region during intervals and during the whole transmission. Furthermore, the ISP can detect the popularity of individual sessions and derive a rough estimate of regions which were densely populated during the whole period of transmission.

The next section presents the benefits of using our scheme.

5 BENEFITS OF OUR SCHEME

The following are the advantages we claim of using our scheme in large RTCP groups:

- *Resolving the storage scalability problem:* Members do not store state about every distinct member in the group because they do not need to know the group size.

- *Timely reporting of feedback reports:* Feedback reports become more useful because the RTCP reporting interval does not depend on the group size so feedback delay is minimised. Hence, the experience of members during short intervals of the whole transmission is accurately reported.
- *Effective use of the bandwidth:* Using our scheme, the number of incoming RRs to every member in the group is decreased and there is limited travelling of RRs. This is because of the formation of local regions where RRs are not multicast but are sent with limited scope and not global scope.
- *Decrease in the number of redundant reports:* Even though, in every region we can still have redundant RRs sent to the AG of the region, the total number of redundant RRs, which used to be multicast, is decreased. In addition, measurements in RRs are aggregated into AGRs summarising the quality of the received data.
- *Useful statistics to be used in network diagnosis and in charging:* The aggregated statistics received by the Manager can help a network administrator to diagnose problems in the network. In addition, these QoS measurements can help an Internet Service Provider (ISP) to estimate the quality received in certain regions and the total number of members in the group can show the popularity of individual sessions.

6 SUMMARY AND FUTURE WORK

We have presented the problems encountered in the deployment of RTCP in large dynamic groups. Also, we discussed the functionality of RTCP Receiver Reports (RRs) in the current Internet where lots of adaptive applications are using receiver-based rate-adaptive schemes instead of schemes based on sender adaptations. We have designed a Scalable RTCP (S-RTCP) scheme in which members are organised dynamically in local regions; every region has an Aggregator (AG) that receives RRs locally from its members, extracts useful information, derives some statistics, then sends this information to a Manager. The Manager monitors the quality of the data distribution and performs some statistical analysis to estimate which regions are suffering from congestion. We believe our scheme reduces some of the RTCP scalability problems encountered in large groups, namely feedback delay, increase in storage state, and ineffective use of the RTCP bandwidth especially for receivers connected through low bandwidth links. In addition, our scheme directs important information included in RRs to an entity that can make valuable use of them.

In the next phase of our work, we are simulating S-RTCP using the network simulator (NS) (McCanne, 1998) and we will report the results in due course. Also, we intend to investigate more functions that the Manager can do, analyse the limitations of our design, and try to refine it.

7 ACKNOWLEDGEMENT

We wish to acknowledge the support of BT Labs for sponsorship of Randa El-Marakby's Ph.D. programme.

8 REFERENCES

- Aboba, B. (1996) "Alternatives of Enhancing RTP Scalability", Internet Draft, draft-aboba-rtpscale-02.txt, Nov. 1996.
- Bolot, J. and Turetti, T. (1994) "A rate control mechanism for a packet video in the Internet", Proceedings of IEEE Infocom'94, Toronto, pp. 1216-1223, June 1994.
- Busse, I., Deffner, B. and Schulzrinne, H. (1996) "Dynamic QoS Control of Multimedia Applications based on RTP", Computer Communications, vol. 19, no. 1, pp. 49-58, January 1996.
- El-Marakby, R. and Hutchison, D. (1997a) "Towards Managed Real-time Communications in the Internet Environment", Proceedings of the Fourth IEEE Workshop on the Architecture and Implementation of High Performance Communication Systems (HPCS'97), Greece, June 1997.
- El-Marakby, R. and Hutchison, D. (1997b) "Delivery of Real-time Continuous Media over the Internet", Proc. of the second IEEE Symposium on Computers and Communications (ISCC'97), Alexandria, Egypt, pp. 22-26, July 1997.
- El-Marakby, R. and Hutchison, D. (1998) "Scalability Improvement of the Real-time Control Protocol (RTCP) Leading to Management Facilities in the Internet", to appear in the Proceedings of the third IEEE Symposium on Computers and Communications (ISCC'98), Athens, Greece, June 1998.
- McCanne, S. and Floyd, S. (McCanne, 1998) "LBNL Network Simulator", <http://mash.cs.berkeley.edu/ns>, 1998 version.
- McCanne, S. and Jacobson, V. (1996) "Receiver-driven Layered Multicast", 1996 ACM Sigcomm Conference, pp. 117-130, August 1996.
- Meyer, D. (1997) "Administratively Scoped IP Multicast", Internet Draft, draft-ietf-mboned-admin-ip-space-03.txt, June 1997.
- Papadopoulos, C., Parulkar, G., Varghese, G. (1998) "An Error Control Scheme for Large-Scale Multicast Applications", Proceedings of Infocom'98, San Francisco, CA, pp. 1188-1196, 1998.
- Rosenberg, J. and Schulzrinne, H. (1997a) "Timer Reconsideration for Enhanced RTP Scalability", draft-ietf-avt-reconsider-00.ps, July 1997.
- Rosenberg, J. and Schulzrinne, H. (1997b) "New Results in RTP Scalability", draft-ietf-avt-byerecon-00.ps, November 1997.
- Schulzrinne, H., Casner, S., Frederick, R. and Jacobson, V. (1996) "RTP: A Transport Protocol for Real-time Applications", RFC 1889, January 1996.
- Schulzrinne, H., Casner, S., Frederick, R. and Jacobson, V. (1997) "RTP: A Transport Protocol for Real-time Applications", draft-ietf-avt-rtp-new-00.ps, December 1997.
- Thaler, D., Aboba, B. (1997) "Multicast Debugging Handbook", Internet Draft, draft-ietf-mboned-mdh-00.txt, March 1997.
- Yavatkar, R., Griffioen, J. and Sudan, M. (1995) "A Reliable Dissemination Protocol for Interactive Collaborative Applications", Proceedings of ACM Multimedia '95, pp. 333-344, 1995.

9 BIOGRAPHY

Randa El-Marakby is a Ph.D. candidate at Lancaster University. Randa received a BSc from the American University in Cairo, Egypt and a MSc from University of North Texas, United States, both in Computer Science, in 1988 and 1993 respectively. She has worked for ICL and KPMG Peat Marwick. She is interested mainly in real-time services and protocols for the Internet.

David Hutchison is Professor of Computing at Lancaster University and has worked in the areas of computer communications and distributed systems for the past 15 years. He has completed many UK and European funded research contracts and published over 100 papers as well as writing and editing books on these areas. The main theme of his current research is architecture, services and protocols for distributed multimedia systems. He is involved in several UK and European collaborative projects, in which an integrating theme is Quality of Service for multimedia communications. He is Honorary Editor of the Distributed Systems Engineering Journal, and is a programme committee member for many international conferences and workshops. He has just finished a year's sabbatical leave as a visiting academic at HP Labs in Bristol, UK, at EPFL in Lausanne, Switzerland, and at BT Labs in Ipswich, UK.