March 2007

End-to-End Delay Measurement for Instant Messaging Relay Nodes

Muhammad T. Alam
Bond University, Muhammad_Alam@bond.edu.au

Zheng da Wu
Bond University, Zheng_Da_Wu@bond.edu.au

Follow this and additional works at: http://epublications.bond.edu.au/infotech_pubs

End-to-End Delay Measurement for Instant Messaging Relay Nodes

Muhammad T. Alam, Student Member *IEEE*, Zheng Da Wu, Member *IEEE*
School of IT
Bond University, Australia
muhamta123@gmail.com

ABSTRACT

In this paper, we provide complete end-to-end delay analyses including the relay nodes for instant messages. Message Session Relay Protocol (MSRP) is used to provide congestion control for large messages in the Instant Messaging (IM) service. Large messages are broken into several chunks. These chunks may traverse through a maximum number of two relay nodes before reaching destination according to the IETF specification of the MSRP relay extensions. We discuss the current solutions of sending large instant messages and introduce a proposal to reduce message flows in the IM service. The analysis presented in this paper is divided into two parts. At the former part, we consider virtual traffic parameter i.e., the relay nodes are stateless non-blocking for scalability purpose. This type of relay node is also assumed to have input rate at constant bit rate. The later part of the analysis considers relay nodes to be blocking and the input parameter to be exponential. The performance analysis with the models introduced in this paper is simple and straight forward, which lead to reduced message flows in the IM service. Also, using our model analysis a delay based optimization problem can be easily deduced.

Keywords: Instant messaging, MSRP, Stateless, Chunking.

1   INTRODUCTION

In this paper, we provide complete end-to-end delay analyses including the relay nodes for instant messages. Instant messaging (IM) is one of today's most popular services. Thus, it is not a surprise that 3G IP Multimedia Subsystem (IMS) already has this service well supported in its architecture. IM is the service that allows an IMS user to send some content to another user in near-real time. The content in an instant message is typically a text message, but can be an HTML page, a picture, a file containing a song, a video clip, or any generic file.

There are two modes of operation of the instant message service, depending on whether they are stand-alone instant message, not having any relation with previous or future instant message. This mode of IM is referred to as "pager mode". The model is also similar to the SMS (Short Message Service) in cellular networks. The other model is referred to as session based instant message that is sent as part of an existing session, typically established with a SIP (Session Initiation Protocol) INVITE request. Both modes have different requirements and constraints, hence the implementation of both models.

The IETF (Internet Engineering Task Force) has created an extension to SIP that allows a SIP UA (User Agent) to send an instant message to another UA. The extension consists of a new SIP method named MESSAGE. The SIP MESSAGE method (RFC 3428 [1]), is able to transport any kind of payload in the body of the message, formatted with an appropriate MIME (Multipurpose Internet Mail Extensions) type. 3GPP TS 23.228 [25] already contains requirements for Application Servers (ASs) to be able to send textual information to an IMS terminal. 3GPP TS 24.229 [2] introduces support for the MESSAGE method extension. The specification mandates IMS terminals to implement the MESSAGE method [1] and to allow implementation to be an optional feature in ASs.

The work over instant messaging [4, 5, 6] observed so far lacks a thorough analysis of the scalable behavior of the nodes involved in providing the IM service. The messages of IM may be very large. Large instant messages have disadvantages like service behavior is too slow on low bandwidth links and more importantly, messages get fragmented over some transport protocol and then look at SIP extension that resolve this issue. Even if messages are compressed, sometimes SIP messages can be too large. Another problem with SIP is that the fact that any proxy can change the transport protocol from TCP (Transmission Control Protocol) to UDP (User Datagram Protocol) or other transport protocols and vice versa. The protocols other than TCP and SCTP (Stream Control Transmission Protocol) are not famous for congestion control. If an IMS terminal is sending a large instant message over a transport protocol that does not offer congestion control, the

network proxies can become congested and stop processing other SIP requests like INVITE, SUBSCRIBE, etc. Even if a terminal sends large SIP MESSAGE over a transport protocol that implements end-to-end congestion control e.g., TCP, SCTP, the next proxy can switch to UDP and congestion may occur.

To solve the issue of large message passing and congestion control in IM, a limit has been placed on the SIP MESSAGE method such that MESSAGE requests cannot exceed the MTU (Maximum Transmit Unit) minus 200 bytes. If the MTU is not known, this limit is 1300 bytes. Another solution to sending SIP MESSAGE requests with large bodies to use the content indirection mechanism [3]. Content indirection allows replacing a MIME body part with an external reference, which is typically an HTTP (Hyper Text Transfer Protocol) URI (Universal Resource Identifier).

Another solution to getting around the size limit problem with MESSAGE is to use session-based IM mode rather than pager mode. Session-based instant message mode uses the SIP INVITE method to establish a session. An IMS terminal establishes a session to send and receive instant messages via Message Session Relay Protocol (MSRP) [24]. MSRP is a simple text-based protocol whose main characteristic is that it runs over transport protocols that offer congestion control. In the IMS, MSRP is implemented in the IMS terminals. Analysis is required to determine the service order of such servers. Our work in this paper is to analyze the delay bound of the relay nodes that implements the MSRP to provide instant messaging service. The benefit of the work lies in the simplicity of the model derivation.

The rest of the paper is organized as follows. Section 2 provides a review of MSRP. The SEND chunking system, our proposal of scalability over MSRP relay nodes, delay analysis for both work conserving non-blocking and blocking situation are described in Section 3. Finally we conclude the paper in Section 4.

2 BACKGROUND

There are currently three methods defined in MSRP after the INVITE message is sent for an IM session set up: i) SEND: sends an instant message of any arbitrary length from one endpoint to another, ii) VISIT: and endpoint connects to another end point, and iii) REPORT: endpoint or a relay provides message delivery notifications.

MSRP does not impose any restriction on the size of an instant message. If an IMS user, Alice wants to deliver a very large message, she can split the message into chunks and deliver each chunk in a separate SEND request. The message ID corresponds

to the whole message, so the receiver can also use it to reassemble the message and tell which chunks belong with which message.

Long chunks may be interrupted in mid-transmission to ensure fairness across shared transport connections. This chunking mechanism allows a sender to interrupt a chunk part of the way through sending it. The ability to interrupt messages allows multiple sessions to share a TCP connection, and for large messages to be sent efficiently while not blocking other messages that share the same connection, or even the same MSRP session. Any chunk that is larger than 2048 octets MUST be interruptible [24].

Another characteristic of MSRP is that, MSRP messages no not traverse SIP proxies. This is an advantage, since SIP proxies are not bothered with proxying large instant messages. Also, MSRP does not run over UDP or any other transport protocol that does not offer end-to-end congestion control. It supports instant messages to traverse zero, one or two MSRP relays (see Figure 1). The relay extension of MSRP is defined in [7].
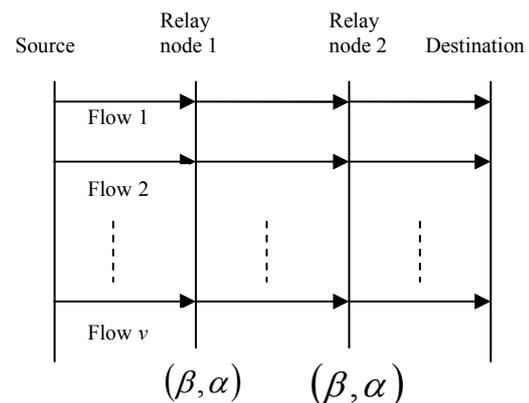

Figure 1: IM with maximum 2 relay nodes

The default is that SEND messages are acknowledged hop-by-hop. Each relay node that receives a SEND request acknowledges receipt of the request before forwarding the content to the next relay or the final target. When sending large content, the client may split up a message into smaller pieces; each SEND request might contain only a portion of the complete message. For example, when Alice sends Bob a 4GB file called "file.mpeg", she sends several SEND requests each with a portion of the complete message. Relays can repack message fragments en-route. As individual parts of the complete message arrive at the final destination client, the receiving client can optionally send REPORT requests indicating delivery status. MSRP nodes can send individual portions of a complete message in multiple SEND requests. As relays

receive chunks they can reassemble or re-fragment them as long as they resend the resulting chunks in order.

A series of papers [8-10] have studied the capacity scaling in relay networks. These works quantify the impact of large wireless relay networks in terms of signal-to-noise ratio. Most of the work focuses on characterizing one relay node only. The work of H. Bolcskei *et all* in [10] demonstrated that significance performance gains can be obtained in wireless relay networks employing terminals with multiple-input multiple-output (MIMO) capability. However, these works do not address the issue of characterizing traffic parameter in relay nodes where the relay nodes do not keep the transaction states. A signification challenge is to schedule the large chunks and characterize the traffic parameters under delay bounds.

In any IMS network the capacity (memory/storage) is large for IM communication. Large messages have to be broken down into chunks to overcome the fixed size limit fact. Real time service of IM is always desirable. However, issues arise if a) the relay nodes in between source and destination IMS terminals possess slow links b) traffic order gets distorted before reaching relay nodes and c) relay nodes maintains transaction states. Therefore efficient service discipline of the chunks of IM is necessary. In an IM system, transmission time typically depends on the number of chunks in messages. Moreover, service time of the chunks depends on batch size arrivals. The number of broken chunks in a large message is not fixed. Thus analysis of such system is not trivial. In this work, we explore delay characteristics of instant messages when the messages traverse via relay nodes.

The end to end delay bound of IMS instant messages indeed requires much attention. Although, the study of the fundamental frameworks, namely Integrated and Differentiated services have a long history, defining the flow characteristic of an IMS instant message traversing the relay nodes (maximum number of relay nodes is two for an IMS terminal [7]) under MSRP is not trivial due to the arbitrary number of chunks in IM messages. We analyze the end-to-end delay for an IM under work conserving situation. The delay bound is useful to formulate the optimization of transmission IM end-to-end transmission time.

3 MODELLING

The large sized SEND messages in IM, MSRP delivers in several SEND messages, where each SEND contains one chunk of the overall message. The crucial aspect in this paper is the ordering of SEND chunks at the relay nodes if they do not keep transaction states of a chunk flow for scalability purpose.

Long chunks are interrupted in mid-transmission to ensure fairness across shared transport connections. To support this, MSRP uses a boundary-based framing mechanism. The start line of an MSRP request contains a unique identifier that is also used to indicate the end of the request. Included at the end of the end-line, there is a flag that indicates whether this is the last chunk of data for this message or whether the message will be continued in a subsequent chunk. There is also a Byte-Range header field in the request that indicates that the overall position of this chunk inside the complete message.

This chunking mechanism allows a sender to interrupt a chunk part of the way through sending it. The ability to interrupt messages allows multiple sessions to share a TCP connection, and for large messages to be sent efficiently while not blocking other messages that share the same connection, or even the same MSRP session. As mentioned before that any chunk that is larger than 2048 octets MUST be interruptible. While MSRP would be simpler to implement if each MSRP session used its own TCP connection, there are compelling reasons to conserve connection. For example, the TCP peer may be a relay device that connects to many other peers. Such a device will scale better if each peer does not create a large number of connections. The chunking mechanism only applies to the SEND method, as it is the only method used to transfer message content [24]. We call the chunking mechanism i.e., breaking one large SEND message into several SEND messages a SEND system.

Proposal: Traditional MSRP [24] may be used without traditional session set up in IMS to provide the congestion control. Also, MSRP relay nodes should not keep transaction states for the SEND chunks.

The benefit of this proposed technique contains reduced message flows in the network as well as gaining scalability at the relay nodes. In order to comply with this, we propose the following scheduling. The detail analysis is provided below that captures the delay bound of the relay nodes.

We provide analysis of one relay node first in terms of aggregate flows of two SEND message chunk flows and delay bound, which will later be used to compute delay bound for SEND systems with two relay nodes, source and destination. We assume the following for the analysis. One IMS source terminal sends multiple large instant messages (SEND) to the same destination via two relay nodes. Each of the SEND messages is broken into small SEND chunks. It is to be noted that we are assuming relay nodes do not keep transaction states of the chunks. Though, the IETF draft [7] specifies that the relays may keep transaction states for a very short time, it will be expensive to keep such states

for the relay nodes if there is huge number of clients being served and traffic flows are massive. We are only interested in busy traffic situation for the derivation of properties in this research and thus server analysis with stateless assumption is more practical. We assume message chunks are served according to the order of delivery time tag of the previous node and all chunks are treated as if they belong to a single flow due to the elastic and massive flows of SEND chunk messages in a flow. Thus performance analysis of an individual flow at a relay node can be achieved by analyzing the aggregated flows at this node.

3.1 For Non-blocking Relay Nodes

Here we adopt a scheduler which services a job according to delivery time stamps of the pervious node. We aim to provide work conserving but stateless scheduling of the chunks. Every message chunks has message id that identifies which SEND message it belongs. The source node generates the chunks and delivers them to a relay node. The ordering is considered to be the order as the source node generates chunks for the first relay node and then first relay node for the second relay node and so on. During the delivery time these chunks may receive time stamp tags. These chunks may reach / propagate to the relay node out of order, and hence the arrival times of the chunks to the relay node may not always be in order of the order id of the message chunks.

Let the propagation delay and link capacity of any link are 0 and c, respectively. The sequence of chunks transmitted by a source to a destination is referred to as a flow (Figure 1). The paths via relay nodes are predetermined as defined in the MSRP relay extensions [7] (The relay nodes are authorized by explicitly by the end terminals). Let, at a relay node chunk k of a flow i is attached with a time stamp tag according to the delivery time from the previous node of $A_i^k + \left(\dfrac{L_i}{\alpha_i} \right)$ where α_j, L_i, and

A_i^k are the input rate, chunk size and the arrival time of chunk k of flow i respectively; the delivery order time stamp of chunk k of flow i is updated at the next relay node with an increment of $\left(\dfrac{L_{max}}{c} \right) + \left(\dfrac{L_i}{\alpha_i} \right)$, and chunks are served at the increment order of their previous node's delivery order time tag, where, L_{max} is the maximum size of chunk in all flows. Under these conditions, it is easy to perceive that the worst case delay of a flow i at

any relay node is no longer than $\left(\dfrac{L_{max}}{c} \right) + \left(\dfrac{L_i}{\alpha_i} \right)$ [11].

We adopt the characteristic of traffic model in [12, 13] which has been widely adopted for characterizing network traffic. If the total traffic of a flow $F(t_1, t_2)$ arriving in the time interval $(t_1, t_2]$ is bounded by:

$$F(t_1, t_2) \le \sigma + \rho(t_2 - t_1) \qquad (1)$$

Then the flow is referred to as conforming to the traffic parameter (σ, ρ). Here the assumptions are under non-overflow condition with a flow injection to a leaky bucket with parameters of buffer size, σ and output rate ρ. In other words, ρ is the average traffic rate in the long run and σ is the burst bound of the flow (σ, ρ). It is practical to assume that the links of relay nodes will be subject to delay bound in terms of propagation delay. We consider a chunk to be arrived only after its last bit has arrived to a relay node and the delivery time of a chunk at a node is the time when the last bit of the chunk leaves the relay node. Note that we are considering the input traffic as the constant bit rate for the relay nodes in this section.

If we consider steady state of the network i.e., traffic load less than one then a chunk will only be delayed at a node if there is a chunk being served or there are chunks waiting in the buffer with earlier delivery time stamps, we assume that the start time of each busy period is initialized at 0. Here, a busy period is an interval of time during which the transmission queue of the output link is continuously backlogged which is consistent with [12].

The previous node's delivery time stamp tag of each chunk lags behind its arrival time at any relay node. Note that chunks are served by the order of their previous node's delivery time stamps which is our assumptions. Thus, the delay for each chunk to traverse the network remains the same only if the time stamps of all chunks are increased by a constant D at the previous node. We can assume that, if the burst of each flow is bounded and the capacity of any link is no less than the average rate of the flows traversing the link, there exists a worst case delay bound in the network, i.e., the worst case delay of a flow to traverse any pair of relay nodes is bounded.

Since we consider that the relay node need not keep transaction states of the SEND flows, chunks in the buffer are served by the order of their delivery time stamp tags, not their arrival times. There is also no distinct relation between the delivery time stamp of a chunk and its arrival time. Thus, a chunk with an earlier delivery time stamp than another chunk, though it arrives later, may be served first. This may happen due to the well-known traffic distortion

problem [12]. Therefore in this regard, it is more reasonable to evaluate a chunk's delay with reference to its previous delivery node's time stamp, rather than its arrival time at the current node. We need to characterize traffic of this kind of scheduling for the relay nodes. We define a parameter (β, α) such that the total traffic of the flow of chunks, whose time stamps are in the range of $(t_1, t_2]$, is no larger than $\beta + \alpha(t_2 - t_1)$, which is similar to the (σ, ρ) traffic model. We assume that chunks are ordered by their previous node's delivery time stamps as $P_1, P_2, ..., P_k, ...(R_i \geq R_j, if\ i > j$, where R_i is the previous node's delivery time tag of chunk P_i). For any two chunks P_m and P_k, $(k \geq m),\ \beta + \alpha(R_k - R_m) \geq \sum_{i=m}^{k} L_i$, where L_i is the size of P_i. A chunk may receive service as long as there is no chunk in the buffer when it arrives. Thus, it is necessary to take into account the arrival time of a chunk to characterize traffic in a relay node. Therefore, we define the traffic parameter for any two chunks of a flow as follows: for any two chunks P_k and P_m of a flow $(k \geq m \geq 1)$,

$$\beta + \alpha(R_k - \max\{R_{m-1}, \min\{A_m, A_{m+1}, ..., A_k\}\}) \geq \sum_{i=m}^{k} L_i$$

where A_i is the arrival time of chunk P_i, $i=1,2,...$; we refer to $F(t_1, t_2) = \beta + \alpha(t_2 - t_1)$ in the time interval $(t_1, t_2]$ as the traffic function of this flow with the traffic parameter (β, α). We apply the additive property of (σ, ρ) traffic model [12] to obtain the following:

Proposition 1: Given two flows with traffic parameters (β_1, α_1) and (β_2, α_2) the traffic parameter of the aggregated traffic of the two flows is $(\beta_1 + \beta_2,\ \alpha_1 + \alpha_2)$.

Proof: Assume that chunks are ordered by their delivery order. Given any two chunks P_k and $P_m (k \geq m)$ of the aggregated flow, assuming chunks $P_{i_1},\ P_{i_2}, ...$ and $P_{i_n},\ (i_1 < i_2 < .... < i_n\ and\ n \leq (k-m+1))$ belong to flow 1, and the rest of the chunks $P_{j_1},\ P_{j_2}, ...$ and $P_{j_p},\ (j_1 < j_2 < .... < j_p\ and\ p \leq (k-m+1))$ belong to flow 2. Thus for the virtual traffic parameter, we have

$$\beta_1 + \alpha_1\left(R_{i_n} - \max\{R_{i_1-1},\ \min\{A_{i_1},\ A_{i_2}, ..., A_{i_n},\}\}\right)$$

$$\geq \sum_{s=i_i}^{i_n} L_s$$

$$\hspace{6cm} (2)$$

$$\beta_2 + \alpha_2\left(R_{j_p} - \max\{R_{j_1-1},\ \min\{A_{j_1},\ A_{j_2}, ..., A_{j_p},\}\}\right)$$

$$\geq \sum_{s=j_1}^{j_p} L_s$$

$$\hspace{6cm} (3)$$

Since,

$$\max\{R_{i_n},\ R_{j_p}\} = R_k,$$

$$\min\left\{\begin{array}{l} \min\{A_{i_1},\ A_{i_2}, ..., A_{i_n},\}, \\ \min\{A_{j_1},\ A_{j_2}, ..., A_{j_p},\} \end{array}\right\}$$

$$= \min\{A_m,\ A_{m+1}, ..., A_k\}\ and$$

$$\min\{R_{i_1-1},\ R_{j_1-1}\} = R_{m-1},$$

We have,

$$(\beta_1 + \beta_2) + (\alpha_1 + \alpha_2)$$

$$\times(R_k - \max\{R_{m-1},\ \min\{A_m, A_{m+1}, ..., A_k\}\})$$

$$\geq [\beta_1 + \alpha_1 \times (R_{i_n} - \max\{R_{i_1-1},\ \min\{A_{i_1},\ A_{i_2}, ..., A_{i_n}\}\})]$$

$$+ [\beta_2 + \alpha_2 \times (R_{j_p} - \max\{R_{j_1-1},\ \min\{A_{j_1},\ A_{j_2}, ..., A_{j_p}\}\})] \geq \sum_{s=m}^{k} L_s$$

$$\hspace{6cm} (4)$$

Application of Theorem 1: If the function of all traffic flows are known, the virtual traffic aggregated function can be derived by Theorem 1.
However, the chunk pattern may be distorted at a relay node. In such case, we can provide the following relation for a flow in terms of worst case delay of the outgoing traffic.

Proposition 2: Assume that the traffic parameter of the input traffic of a SEND chunk flow at a relay node is $(\beta,\ \alpha)$ and the worst case delay to traverse a relay node is D (let the mean service time of a chunk at this current node is d). We can characterize the output traffic of this flow as $(\beta',\ \alpha)$ where the buffer requirement is

$$\beta' =\ \max\{0,\ \alpha(D-d) + L_{max}\} + \beta.$$

Proof: Assume that chunks are ordered by their delivery times at this current node, i.e., for chunks P_k and $P_m (k \geq m,\ T_k \geq T_m)$ where the delivery order time tag of chunk P_i, $i = 1, 2, ...,$ is T_i and is also the arrival time of P_i of the output traffic. As

the worst case delay of a chunk is *D,* we have the following relation:

$$T_i \leq R_i + D \qquad (5)$$

Again, since the delivery order of each chunk is delayed by *d* and $\beta' = \max\{0, \ \alpha(D-d)+L_{\max}\}+\beta.$, for any two chunks *k* and *m* $(k \geq m \geq 1)$, we get

$$\beta' + \alpha[R_k + d - \max\{R_{m-1}+d, \ T_m\}]$$
$$\geq \beta' + \alpha[R_k + d - \max\{R_{m-1}+d, \ R_m + D\}]$$
$$\geq \min\ \{\beta + \alpha(R_k - R_{m-1}), \ \beta + L_{\max} + \alpha(R_k - R_m)\} \qquad (6)$$

And

$$\beta + \alpha[R_k - \max\{R_{m-1}, \min[A_m, A_{m+1}, ..., A_k]\}] \ \geq \sum_{i=m}^{k} L_i$$

i.e., $\alpha(R_k - R_{m-1}) \geq \sum_{i=m}^{k} L_i$

$$(7)$$

Now let the previous node's delivery order of a chunk P_i, $i = 1,2,...$, at the outgoing link of the relay node is: $R'_i = R_i + D + \left(\dfrac{L_{\max}}{\alpha}\right)$. Thus from Eq. (6) and (7) we have:

$$\beta + \alpha[R'_k - \max\{R'_{m-1}, \min[T_m, T_{m+1}, ...T_k]\}]$$
$$\geq \min\{\beta + \alpha(R_k - R_{m-1}), \beta + L_{\max} + \alpha(R_k - R_m)\}$$
$$\geq \min\left\{\sum_{i=m}^{k} L_i, \ \sum_{i=m+1}^{k} L_i + L_{\max}\right\} \geq \sum_{i=m}^{k} L_i$$

$$(8)$$

Thus the characteristic of traffic parameter for worst case Delay *D* is $(\beta', \ \alpha)$ and proposition 2 is proved.

Next we analyse the worst case delay bound of a SEND chunk flow to traverse a relay node.

Proposition 3: Let, P_k^i be the *k*th chunk of flow *i* and assume that the chunks are ordered according to their current node's delivery order time tag. Define $\theta_i = \max\{\min_{k \geq m > 1} \{R_{m-1}^i - \min\{A_m^i, A_{m+1}^i, ..., A_k^i\}\}, 0\}$ where R_m^i and A_m^i are the delivery time tag from previous node and the arrival time at current node of P_m^i; $L_{\max}$ be the maximum size of a chunk. Assume that the input traffic of a relay node consists of flows 1,2,..., v , whose traffic parameters are (β_i, α_i) respectively and the capacity of the output link of

this node is $c, c \geq \sum_{i=1}^{v} \alpha_i$. Under these assumptions, the wost case delay bound at a current relay node is:

$$\frac{1}{c}\left[\sum_{i=1}^{v}(\beta_i - \alpha_i\theta_i) + L_{\max}\right] \qquad (9)$$

Proof: For any chunk P_k if we assume *m* to be the biggest integer $k > m > 0$ such that $R_k < R_m$ and $T_k > T_m$ where R_i and T_i are the previous node's delivery time tag and the delivery time of P_i at current node. Thus

$$R_m > R_k \geq R_i, \qquad \text{for all} \qquad m < i < k \qquad (10)$$

$$T_k > T_i \geq T_m, \qquad \text{for all} \qquad m < i < k \qquad (11)$$

In other words, P_m is transmitted before chunks $P_{m+1}, ..., P_k$; however, its previous node's delivery time tag is greater than that of chunks $P_{m+1}, ...P_k$. Thus

$$\min\{A_{m+1}, ..., A_k\} > T_m - \frac{L_m}{c} \qquad (12)$$

Since, $P_{m+1}, ..., P_k$ arrive after $T_m - \left(\dfrac{L_m}{c}\right)$ and depart before P_k at the current relay node, we have

$$T_k = T_m + \frac{\sum_{i=m+1}^{k} L_i}{c} \qquad (13)$$

Note that $R_i \geq A_i$ *for all* $i = 1,2,...,$ and thus

$$R_k \geq R_i \geq A_i \geq T_m - \left(\frac{L_m}{c}\right) \qquad \text{for}$$

$i = m+1, ..., k-1$. Furthermore we have the traffic function,

$$\theta_i = \max\{\min_{k \geq m > 1}\{R_{m-1}^i - \min\{A_m^i, A_{m+1}^i, ...A_k^i\}\}, 0\}$$
$$= \min\{A_m^i, A_{m+1}^i, ...A_k^i\} + \theta$$
$$\leq \max\{R_{m-1}^i, \min\{A_m^i, A_{m+1}^i, ...A_k^i\}\}$$

i.e.,

$$\beta_i + \alpha_i[R_k^i - (\min\{A_m^i, A_{m+1}^i, ...A_k^i\} + \theta_i)] \geq \sum_{j=1}^{k} L_j^i \qquad (14)$$

Since, chunks $P_{m+1}, ..., P_k$ comprise the chunks of flows 1,2,... v , we have

$$\sum_{i=1}^{v}\left\{\beta_i+\alpha_i\left[R_k^i-\left(\min\left\{A_m^i,A_{m+1}^i,...,A_k^i\right\}+\theta_i\right)\right]\geq\sum_{i=m+1}^{k}L_i\right\}$$

i.e.,

$$\sum_{i=m+1}^{k}L_i\leq\sum_{i=1}^{v}(\beta_i-\alpha_i\theta_i)+\left(\sum_{i=1}^{v}\alpha_i\right)\left[R_k-\left(T_m-\frac{L_m}{c}\right)\right]$$

(15)

From Eq. (13) and Eq. (15) we have

$$T_k=T_m+\frac{\sum_{i=m+1}^{k}L_m}{c}$$

$$\leq T_m+\frac{\left(\sum_{i=1}^{v}\alpha_i\right)\left[R_k-\left(T_m-\frac{L_m}{c}\right)\right]+\sum_{i=1}^{v}(\beta_i-\alpha_i\theta_i)}{c}$$

$$\leq R_k+\frac{L_{\max}}{c}+\frac{\sum_{i=1}^{v}(\beta_i-\alpha_i\theta_i)}{c}$$

(16)

If there does not exist such m, then $P_1,...,P_{k-1}$ all leave the node before P_k and thus have

$$T_k=\frac{\sum_{i=1}^{k}L_i}{c}\leq\frac{\left(\sum_{i=1}^{v}\alpha_i\right)R_k+\sum_{i=1}^{v}(\beta_i-\alpha_i\theta_i)}{c}$$

i.e.,

$$T_k-R_k\leq\frac{\sum_{i=1}^{v}(\beta_i-\alpha_i\theta_i)}{c}$$

(17)

Thus the delay is bounded by $\dfrac{\sum_{i=1}^{v}(\beta_i-\alpha_i\theta_i)}{c}+\dfrac{L_{\max}}{c}$ and proposition 3 is proved.

Application of proposition 2 and proposition 3: The proposed propositions are straight forward for performance analysis. From the above relation, we can also characterize the outgoing traffic parameter of a relay node for a given propagation delay, δ. Let δ be the propagation delay of a chunk of a flow $(\beta,\ \alpha)$ i.e, the propagation delay of a chunk from a relay node to the next relay node. Then the worst case delay of a flow is $D+\delta$ if this is the first relay node i.e., there is no update at the previous node of this flow. Here we assume that this is the first relay node and the flows arrive from the source directly to this node. In this case if all of the chunks of the flow

are updated / serviced by an increment d at the relay node, then input traffic parameter for the next relay node is $(\beta',\ \alpha)$ where buffer requirement is $\beta'=\max\{0,\ \alpha(D+\delta-d)+L_{\max}\}+\beta$.

The delay bound of proposition 3 can further be tightened. For instance, if $\left(\sum_{i=1}^{v}\dfrac{\alpha_i}{c}\right)\to 0$, then the worst case delay bound would be $\left(\sum_{i=1}^{v}\beta_i+L_{\max}\right)\Big/c$. On the other hand, if $\theta=\min_i\{\theta_i\}$, and the delivery time tag at the pervious node of all chunks are decreased by θ, then the traffic functions of all flows remain the same and the actual worst case delay bound from proposition 3 is $\dfrac{\left(\sum_{i=1}^{v}\beta_i+L_{\max}\right)}{c}-\theta$. Therefore, it is possible to tighten the worst case delay as well in this instance. If all chunks' delivery time stamps at the previous node are increased or decreased by a constant at the entrance to a relay node, their delivery time remains unchanged. If all chunks' previous node's delivery time tag decreased by θ, applying proposition 3, for any chunk P_k we have the following:

$$T_k-(R_k-\theta)\leq\frac{\sum_{i-1}^{v}[\beta_i-\alpha_i(\theta_i-\theta)]+L_{\max}}{c}$$

i.e.,

$$T_k-R_k\leq\frac{\sum_{i-1}^{v}[\beta_i-\alpha_i(\theta_i-\theta)]+L_{\max}}{c}-\theta$$

(18)

i.e., the worst case delay is bounded by

$$\frac{\sum_{i-1}^{v}[\beta_i-\alpha_i(\theta_i-\theta)]+L_{\max}}{c}-\theta$$

Now if we take the propagation delay into account, the increment for flow n, $1\leq n\leq v$, should be

$$\frac{\sum_{i-1}^{v}[\beta_i-\alpha_i(\theta_i-\theta)]+L_{\max}}{c}-\theta+\delta_{n,i}\qquad\text{where}$$

$\delta_{n,i}$ is the propagation delay of flow n to traverse the link between relay node i and its next adjacent relay node.

Example: In order to further analyze the

proposed propositions, consider two cases.

Let two flows; flow 1 and flow 2 are contending for the bandwidth of a link with a capacity of $\frac{2L}{c}$. The reserved bandwidths of the two flows are both $\frac{L}{c}$, and all chunks are of size L. However, the inter-arrival times of two consecutive chunks of flows 1 and 2 are c and $c/2$, respectively. Assume that the first chunks of both flows arrive at time 0, and the arrival time of the kth chunk of flow i, i=1,2, is A_i^k, where $A_i^k = (k-1)c$ if i=1, and $A_i^k = \frac{(k-1)c}{2}$ if i=2. The previous node's time tag attached to the kth chunk of flow i is, however, kc, which is independent of i and will make each flow attain its reserved bandwidth. Therefore, it can be observed that the worst case delay of flow 1 is c, and it is infinity for flow 2. However, if the previous node's time tag of the kth chunk of flow i, i=1,2, is set to $A_i^k + c$, then the worst case delays of both flows become infinity. We can observe such characteristic from the propositions we derived. The delivery order at the previous node attached to the kth chunk of flow 2 are i) kc and ii) $A_i^k + c$ respectively. In the first case the traffic parameters of the two flows are $\left(0, \frac{L}{c}\right)$ and $\left(0, \frac{L}{c}\right)$ i.e., same. By the aggregate property from proposition 1, we have the traffic parameter of the aggregate flow as $\left(0, \frac{2L}{c}\right)$ and by proposition 3 the delay bound of any chunk is $\frac{c}{2}$ since $\theta_1 = \theta_1 = 0$. Therefore, since the delivery order at the previous node of a chunk lags behind its arrival time, bounded by c and infinity in flows 1 and 2 respectively, and then the worst case delay of the flows are $\frac{3c}{2}$ and infinity respectively. In the later case, the traffic parameter of flow 2 is $\left(\infty, \frac{L}{c}\right)$.

Thus the aggregate traffic flow is $\left(\infty, \frac{2L}{c}\right)$ and the worst case delay is infinity. Thus, we see that the worst case delays of both flows become infinity according to our analysis.

From the above analysis using proposition 3, we can find the end-to-end delay bound for an IMS source to an IMS destination terminal using two relay nodes as:

$$\sum_{r=1,2}\left[\frac{\sum_{i=1}^{v}(\beta_{ri} - \alpha_{ri}\theta_i) + L_{\max}}{c}\right] + \sum_{i=1}^{v}\alpha_{2i}\delta_{2,e}$$

(19)

Where, r represents the index of relay nodes, α_{2i} is the traffic rate of flow i reaching the end destination terminal e from the second relay node, and $\delta_{2,e}$ is the propagation delay for a chunk of a flow to reach from the second relay node to the destination end terminal. Eq. (19) achieves the goal of our work in stateless work conserving situation. Note that the design and analysis of the above work are consistent with [18-20] with the traffic parameter behaving as virtual clock arrivals as shown in [21]. Li and Knightly [22, 23] provided analysis for multihop stateless scheduling, but the simplicity of our analysis is perhaps preferred to be deployed regarding virtual traffic flows.

So we far we developed scheduling model with consideration that the relay nodes are in steady state that is the traffic load is less than one. However, this situation may not hold and the chunks may be blocked in overload condition. Losses may occur because chunks are rejected when they arrive at a full (arrival) buffer. In this case a full retransmission is initiated by the sender IMS terminal after a timeout, a significant increase in the end-to-end delay. Chunks may also be corrupted as they arrive a relay node. There is a large series of recent work on the asymptotic analysis of loss messages with different server characteristics. The characteristic of loss probability for corrupted messages can be located in Abramov's PhD work in [15]. Abramov showed the effect of adding redundant chunks to a large message. Summarizing from [14, 15] to our context, adding redundant chunks when the load at relay node is slightly greater than 1, will decrease message loss probability with the rate of geometric progression. But, adding redundant chunk is not profitable if the load is much greater than 1, i.e., the loss probability will increase for this case.

3.2 For Overloaded Relay Nodes

In this section of the paper, we analyze chunks being blocked or rejected due to full buffer. In the previous section we presented delay analysis for one source and destination with two intermediate relay nodes. In practical, there may be times chunks arrive in batches at the relay nodes from the same IMS terminal [7]. Alternatively, chunks from the different sources can traverse the same two relay nodes in a network. We focus on analyzing the response time

and mean transmission times of chunks in batch arrival conditions in this section. Let the batches of chunks arrive as Poisson process with mean arrival rate λ_0 (number of *batch* arrivals per unit time) and mean batch size b. We provide thorough derivation of mean transmission time from a source IMS terminal to a destination IMS terminal including two relay nodes in between them. Here we focus on analyzing the response time and mean transmission times of SEND systems.

In a batch-Poisson stream, we can assume that successive batches arrive after intervals which are independent and exponentially distributed. Let the service rates (service times being independent of each other and of the arrival process), are μ_1 and μ_2 for relay node 1 and 2, respectively. The departure process from the first relay node is approximated as Poisson with unit batches. We see that both the relay nodes behave as M/GI/1 servers under the above assumptions though a deeper analysis to correlate the arrival process of the second relay node with the departure process at the first relay node is required. Any chunk which has not transferred successfully is lost. We are interested of the losses due to the full buffer in this section. By full buffer we mean that the relay nodes are considered to have finite buffer only.

We assume the effect on the sender and receiver is the same as a full-buffer loss. Obviously, losses due to a full buffer at each node, cause an additional transmission delay by a timeout of duration much greater than the per-link transmission delay and typically a multiple of the estimated end-to-end delay. Let, the retransmission rate i.e., the net rate at which chunks are resent to the sender node due to losses of all types is λ_r. The total rate of chunk arrivals, including retransmissions, is therefore

$$\lambda = b\lambda_0 + \lambda_r \qquad (20)$$

Let, T_1, T_2, T_3 be the constant transmission delays for a chunk passing over the links between the sender and first relay node, first relay node and second relay node, and second relay node and the receiver IMS terminal respectively. Hence the transmission delay for successful chunks, excluding the time spent in the nodes, is

$$T = T_1 + T_2 + T_3 \qquad (21)$$

If a transmission has been successful, then the receiver sends a REPORT message for the successfully delivered chunk to the sender node through the route, according to the MSRP. When the arrival buffer at a node is full, the rejection of an arriving chunk will result in no 200OK being sent. The value of the sender's timer will reach the pre-specified time-out value, the chunk will be deemed lost and a retransmission will be attempted with a given probability.

The value of the time-out period is set to a multiple k of the estimated mean transmission time, for some $k \geq 1$ set by the user. Let, the rate of successful transmissions be λ''. This makes the probability that a chunk is successfully transmitted to the receiver $\dfrac{\lambda''}{\lambda}$. Thus the probability that a chunk is rejected at either of the relay nodes due to a full buffer is,

$$p_f = 1 - \frac{\lambda''}{\lambda} \qquad (22)$$

An IMS terminal may also define the probability of retransmission p_r. For $p_r = 0$, there will be no retransmissions. This will impact with higher losses of chunks but lower congestion. For $p_r = 1$, there is always a retransmission attempt and so there are no losses and the number of retransmissions is unlimited. This causes extra load on the network which might result in congestion and hence significantly longer delays.

It is desirable for every system that losses do not exceed a specified maximum loss rate L_r. This leads to the relation with the chunk arrival rates as follows:

$$(1 - p_r)(\lambda - \lambda'') \leq b\lambda_0 L_r \qquad (23)$$

We now provide the expressions for relay node utilization below.

For relay node 1, the utilization is

$$U_1 = \frac{\lambda'}{\mu_1} \qquad (24)$$

Where λ' is the throughput from the first relay node. And for the second relay node, the utilization is:

$$U_2 = \frac{\lambda''}{\mu_2} \qquad (25)$$

Note that we assumed λ the batch arrival rates at the first relay node, λ' the batch arrival rates at the second relay node which is the throughput of the first relay node and λ'' the throughput of the second relay node. Here we find that the mean rate of retransmission in the system is,

$$\lambda_r = p_r(\lambda - \lambda'') \qquad (26)$$

With these analyses, it is plain to compute the two throughputs of the relay nodes. For M/M/1 machines performance evaluation can be performed from the following expressions:

$$\lambda' = \frac{\rho_1(1 - \rho_1^{A_1})\mu_1}{1 - \rho_1^{A_1+1}} \qquad (27)$$

$$\lambda'' = \frac{\rho_2(1 - \rho_2^{A_2})\mu_2}{1 - \rho_2^{A_2+1}} \qquad (28)$$

Where, ρ_1, ρ_2 are load at the first relay node and the second relay node respectively; and A_1, A_2 are the buffer sizes of the first relay node and the second relay node respectively.

Let the service time random variable be S, with mean $s = \dfrac{1}{\mu}$, the batch size random variable be B with mean b, Y_n be the queue length immediately after the last chunk in the nth batch departs and $G_X(z)$ be the generating function of the probability mass function of discrete random variable X. Let U_n denotes the number of chunks that arrive during the service of (all the chunks in) the nth batch. Let $\pi_n(k) = P(Y_n = k)$ for $n \geq 1, k \geq 0$, so that $G_{Y_n}(z) = \sum_{k=0}^{\infty} \pi_n(k) z^k$. At equilibrium, assuming this exists, let $\pi_n(.) \to \pi(.)$ and $G_{Y_n}(.) \to G_Y(.)$ as $n \to \infty$. The random variable B denotes a generic batch size random variable B_n and we use V similarly to denote a generic instance of V_n.

Let the sojourn time, or waiting time, in the queue of the last chunk in a batch – i.e. the sum of the time it spends waiting to start service and its service time – be W.

The Laplace-Stieltjes transform of the response time distribution in such an M/GI/1/∞ queue with batch arrivals can then be shown to be given by:

$$[1 - G_B(H)]W^*(\theta) = (1 - \rho_1')[G_B(S^*(\theta)) - G_B(H)] \tag{29}$$

Where $z = G_B^{-1}(1 - \dfrac{\theta}{\lambda'})$ (30)

$$\theta = \lambda'(1 - G_B(z)) \tag{31}$$

and

$$H = z^{-1} S^*(\theta). \tag{32}$$

W is the average response time. * here denotes data indexed by relay node $i = 1, 2$.

The mean waiting time (and arbitrary higher moments) at relay node 1 can be approximated by applying the recurrence formula of the Laplace-Stieltjes transform of the response time distribution of the M/GI/1/∞ queue with batch arrivals. The computation of higher moments of waiting time of this type can be located in [16, 17].

We have,

$$W_1 = -W^{*\prime}(x)|_{x=0} \tag{33}$$

Where

$$W^{*\prime}(0) = \frac{U_1(2 - U_1)(M_{2B} - b) + (\lambda')^2 b^3 M_{2S}}{2\lambda' b^2(U_1 - 1)} + \frac{U_1(1 - U_1 b)}{\lambda' b(U_1 - 1)} \tag{34}$$

M_{2B} and M_{2S} are the second moments of batch size and service time distributions, respectively.

The total expected response time at relay node 2, W_2 can be computed as the M/M/1 response time approximation as:

$$W_2 = \frac{1}{\mu_2(1 - \rho_2)(1 - \rho_2^{A_2})}\left(1 - (A_2 + 1)\rho_2^{A_2} + A_2\rho_2^{A_2+1}\right) \tag{35}$$

The total expected response time W, i.e., the combined time spent in the two relay nodes on a successful transmission attempt, is the sum of the expected response times at each node, i.e.,

$$W = W_1 + W_2 \tag{36}$$

Thus, the mean transmission time (MTT) for a chunk that is successful on its first attempt is

$$MTT = W + T \tag{37}$$

Eq. (37) achieves the goal of out model. Failed chunks, due to full buffer, retry a number of times given by the retransmission probability p_r. Because each retry is made independently of previous attempts, this number of attempts is a geometric random variable with parameter p_r. The overhead incurred by a failed transmission, i.e. the elapsed time between the start of an attempt that subsequently fails and the start of the next attempt, consists of the time-out delay of $k*MTT$ for chunks lost due to a full buffer (k mean successful transmission times). We express this overhead, L as follows:

$$L = k * MTT * p_f \tag{38}$$

4   CONCLUSIONS

Providing instant messaging in real time is indeed open challenge today. Previous works on relay nodes are centered on one node only. We have shown a complete en-to-end delay evaluation that includes two relay nodes (maximum number that a source MRSP terminal can select) for both buffer blocking and non-blocking situation. In the former analysis constant bit rate was considered to be consistent with Cruz's [12, 13] famous work conserving virtual traffic parameter model (σ, ρ).

With our model, the performance evaluation of end-to-end delay for large instant messages becomes straight forward. The later analysis was provided for buffer overloaded situation. The analysis presented leads to the common optimization problem of minimizing Eq. (37) with the respect to the relay

node throughputs, utilizations and buffer sizes. We believe this is an interesting and useful avenue to be explored for a large SEND system which is our future area of investigation.

5 REFERENCES

[1] B. Campbell, J. Rosenberg, H. Schulzrine, C. Huitema, and D. Gurle, "Session Initiation Protocol (SIP) Extension for Instant Messaging", RFC 3428, Internet Engineering Task Force, (2002).

[2] 3GPP, Internet Protocol (IP) multimedia call control protocol based on Session Initiation Protocol (SIP) and Session Description Protocol (SDP); Stage 3, TS 24.229.

[3] E. Burger, "A Mechanism for Content Indirection in Session Initiation Protocol (SIP) Messages" RFC 4483, Internet Engineering Task Force, (2006).

[4] M. Debbabi, M. Rahman, W. Lin, J. Gopal, S. Sridhar, "Standard SIP-based instant messaging and presence APIs for networked devices", Networked Appliances, Proceedings. IEEE 5th International Workshop on, 30-31 Oct. (2002), pp: 59 - 64

[5] M. Gomez, J. L. Megias, C. Bueno, C. Brocal, "Interworking between the Multimedia Messaging Service (MMS) and the 3G IP Multimedia subsystem (IMS) Instant Messaging Service", IEEE 16th International Symposium on Personal, Indoor and Mobile Radio communications, (2005), pp: 22742-278.

[6] L. Zhou, "An empirical Investigation of Deception Behaviour in Instant Messaging", IEEE Transactions on Personal Communication, Vol: 48 (2), (2005), pp: 147-160.

[7] C. Jennings, R. Mahy, A. B. Roach, "Relay Extensions for the Message Sessions Relay Protocol (MSRP)", Internet Engineering Task force, draft-ietf-simple-msrp-relays-09.txt, (2007), Work on Progress.

[8] M. Gastpar, M. Vetterli, "On the capacity of wireless networks: the relay case" INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, Vol: 3, 23-27 June (2002), pp: 1577 – 1586, Digital Object Identifier 10.1109/INFCOM.2002.1019409

[9] B. Wang, J. Zhang, A. Host-Madsen, "On the capacity of MIMO relay channels", Information Theory, IEEE Transactions on, Vol: 51 (1), Jan. (2005), pp: 29 – 43, Digital Object Identifier 10.1109/TIT.2004.839487

[10] H. B¨olcskei, R. U. Nabar, O. Oyman, A. J. Paulraj, "Capacity Scaling Laws in MIMO Relay Networks", IEEE Transactions on Wireless Communications, Vol: 5 (6), June (2006), pp: 1433-1443.

[11] Z. Zhang, Z. Duan, and Y. Hou, "Fundamental Trade-offs in Aggregate Packet Scheduling", Parallel and Distributed Systems, IEEE Transactions on, Vol: 16 (12) (2005), pp: 1166-1177.

[12] R. L. Cruz, "A calculus for network delay. I. Network elements in isolation", Information Theory, IEEE Transactions on, Vol: 37 (1), (1991), pp: 114 – 131, Digital Object Identifier 10.1109/18.61109.

[13] R. L. Cruz, "A calculus for network delay. II. Network analysis", Information Theory, IEEE Transactions on, Vol: 37 (1), (1991), pp: 132 – 141, Digital Object Identifier 10.1109/18.61110

[14] V. M. Abramov, "Asymptotic behaviour of the number of lost messages" SIAM J. Appl. Math. Vol: 64, (2004), pp: 746-761.

[15] V. M. Abramov, "Asymptotic methods for queuing systems and networks with application to telecommunications", PhD Thesis, School of Mathematics Science, Tel Aviv University, (2004).

[16] N. Gulpinar, P. Harrison, B. Rustem, "Performance Optimization of Mean Response Time in a Tandem Router Network with Batch Arrivals", Network Operations and Management Symposium, 10th IEEE/IFIP, (2006), pp: 1 – 4.

[17] N. Gulpinar, P. Harrison, B. Rustem, "An optimisation model for a two-node router network", Modeling, Analysis, and Simulation of Computer and Telecommunications Systems, 2004. Proceedings IEEE. 4-8 Oct. 2004, pp: 147 – 156.

[18] J. Kaur, H. M. Vin, "Core-stateless guaranteed throughput networks", IEEE INFOCOM, Vol: 3, 2003, pp: 2155 – 2165.

[19] J. Kaur, H. M. Vin, "Core-stateless guaranteed rate scheduling algorithms", IEEE INFOCOM, Vol: 3, 2001, pp: 1484 – 1492.

[20] Z. Zhang, Z. Duan, and Y. Gao , "A core stateless bandwidth broker architecture for scalable support of guaranteed services", Parallel and Distributed Systems, IEEE Transactions on, Vol: 15 (2), 2004, pp: 167 – 182.

[21] Z. Zhang, Z. Duan, and Y. Hou, "Virtual time reference system: a unifying scheduling framework for scalable support of guaranteed services", Selected Areas in Communications, IEEE Journal on, Vol: 18 (12), 2000, pp: 2684 – 2695.

[22] C. Li, E. W. Knightly, "Coordinated multihop scheduling: a framework for end-to-end services", Networking, IEEE/ACM Transactions on, Vol: 10 (6), Dec. 2002, pp: 776 – 789.

[23] C. Li, E. W. Knightly, "Schedulability criterion and performance analysis of coordinated schedulers", Networking, IEEE/ACM Transactions on Vol: 13 (2), April 2005, pp: 276 – 287.

[24] B. Campbell, R. Mahy, C. Jennings, "The Message Session Relay Protocol". Internet-Draft draft-ietf-simple-message-sessions-18, Internet Engineering Task Force, 2007, Work in Progress.

[25] 3GPP, TSG SSA, IP Multimedia Subsystem (IMS) – Stage 2 (Release 7), TS 23.228 v.7.3.0, 2006-03.