MRTP – A Novel Real-time Protocol for Patient-Doctor Online Meetings and Monitoring

Iyad Al Khatib[1, 2, *]

[1] Universitat Politècnica de Catalunya (UPC), Barcelona, Spain
[2] iITC Stockholm, Sweden

Abstract. This paper presents the design of a novel Voice over IP (VoIP) protocol that carries real-time voice and medical data in the same stream while preserving good quality of service (QoS) for online medical care. The work deliberates the reasons for the choice of the Session Initiation Protocol (SIP). The proposed medical real-time protocol (MRTP) is designed to be compatible with the existing Real Time Protocol (RTP) and has advanced features for medical data transmission. The MRTP leverages the real-time performance due to less overhead i.e., less delays. A major challenge is that some medical data is life-critical i.e., losses may cause harm. Hence, reliability and satisfactory QoS are crucial. A major result is the decrease in average-delay by around 74% compared to concurrently utilizing RTP for the audio part and TCP-connections (sockets) for the same medical data.

1 Introduction

The increasing number of connected people over the last two decades brought a higher demand for online applications with cheaper and faster solutions. With the current COVID-19 pandemic, online-conferencing has witnessed an unprecedented growth. For instance, the utilization of the video-conferencing application, Zoom, rose to 300 million participants per day in June 2020 compared to 10 million in December 2019 [1]. Online healthcare applications are not an exception [2]. During the pandemic, this demand soared for distance healthcare consultations [3], where telehealth usage became 38 times higher than what it was before the pandemic and investment in virtual-care and digital-health has skyrocketed reaching a total of USD 14.7 billion [4]. Due to these reasons, the global telehealth market size and share revenue are projected to reach USD 475.5 billion by 2026 [5]. This economic success in providing VoIP services in the new reality depends on reliable communications and user-satisfaction i.e., a satisfied Medical Doctor (MD) and patient.

However, this demand creates QoS problems such as higher delays that may lead to disconnecting users e.g., Zoom totally stopped working worldwide in March 2020 [6]. Such quality problems are expected to affect online healthcare meetings also. This calls for a new cost-effective scalable design for reliable protocols to deploy in telehealth consultations and real-time monitoring. This level of reliability for healthcare monitoring introduces higher technical requirements for QoS parameters such as delay and data/packet loss.

Nonetheless, current technical solutions still use real-time protocols on the TCP/IP suite application-layer and have been the same since decades. There is a mismatch between the current/future medical application needs and the real-time protocols to reduce the number of redundant connections, provide better and faster healthcare, and access larger telehealth resources especially for remote patients. Nowadays, many medical applications and systems send their information over networks with heterogenous components and variable bandwidths i.e., causing higher bandwidth utilization in some connections that face bottlenecks when reaching slower networks on the path. This risks the entire communication especially for critical medical data. The industrial methods to tackle this problem so far focus on purchasing more/better hardware resources i.e., increasing cost. Moreover, the tradeoff between cost and quality for medical applications is arduous due to life-critical data transmission. Hence, revisiting current VoIP protocols to suit such a soaring demand is required.

This paper presents a solution that can utilize the same type of available networks and solve the performance problem via a novel application-layer real-time protocol (MRTP) that aims at enhancing QoS for medical applications. The MRTP is backward compatible with the widely used RTP so that if a user has only an RTP client, the solution can still function i.e., there is no need to invest to change network hardware or software.

Several telehealth solutions require transmitting multiple signals from the same patient to the same medical center/doctor. For example, the 12-lead electrocardiogram (ECG) consists of 12 signals from the same patient, and the devices that store or transmit such signals may fill up to a few GB per patient per day; let alone a large center with many patients or a large online service with hundreds/thousands of patients worldwide. In this respect, network traffic grows rapidly, and the use of relevant protocols and quicker solutions are necessary.

* Corresponding author: iyad.al.khatib@upc.edu

The MRTP is a VoIP real-time protocol that can connect an array of patients (with one or many signals per patients) to distant medical centers/MDs. The challenge is that medical data and audio signals may be concurrent when the patient is in an online meeting with the health-specialist. Hence, audio and datasets are to be sent from the same patient to the same center without loss in performance over heterogenous networks with variant bandwidths. The MRTP considers sending concurrent data from the same patient to the MD via packaging all data chunks (audio and medical data) in one UDP packet i.e., decreasing overhead. For instance, for a the 12-Lead ECG application with an audio call between the MD and patient, one packet will carry all the information (audio and 12 ECG-datasets in one packet) instead of 13 packet streams (each of the 12 ECG-signals on 12 separate streams, and one audio-packet stream). Since UDP does not check for packet loss, the proposed solution utilizes a specifically devised mechanism for checking packet loss on the Application layer (MRTP), which keeps track of MRTP-packet counters (related to dataset-Sequence-numbers, sq_i) and MRTP-acknowledgment signals (MRTP-ACK). Test results show that the decrease in protocol-overhead significantly decreases delay and betters QoS especially for high numbers of patients. When deploying the MRTP, the delay is 74% less than when concurrently utilizing 'RTP for audio-streams and TCP for medical-datasets.'

2 VoIP Protocol Choices

Technically, one reason for the growth in VoIP is that IP networks offer better service flexibility than traditional telephony [7]. However, the challenge is that QoS should remain comparable to that of PSTN [8]. Hence, there is a need for good choices of suitable VoIP protocols that can function together for better QoS. VoIP comprises two types of protocols: signaling (subsection 2.1) and media transmission/reception (subsection 2.2).

2.1. Signaling Protocols

Signaling protocols control the agreement between the caller, callee, and servers to decide how to send voice/media in real-time. There are many VoIP signaling protocols that are deployed in real networks. We ran an investigation to choose the Signaling protocol that best fits our goal of minimizing overhead (to decrease delays). We performed a comparison between different protocols and found that the Session Initiation protocol (SIP) [9-11] is the most appropriate protocol for our purposes for two reasons: it has the fewest number of signaling messages (least overhead), and it is widely deployed. However, for SIP to accept MRTP data, one modification to the header of the SIP Request message (Fig. 1) is needed. Hence, we modified the "Allow" field in the message Header (see the red-colored text in Fig. 1). The "Allow" field is originally designed to include the names of the message-types that are allowed during a SIP-session namely, ACK, BYE, CANCEL, INFO, INVITE, NOTIFY, OPTIONS, and REFER. In our enhanced version, we add the "MRTP" name/type to this list. This keeps SIP Requests backward compatible with SIP clients and servers while differentiating between regular packets that it could carry, and medical data packets that are carried via MRTP.

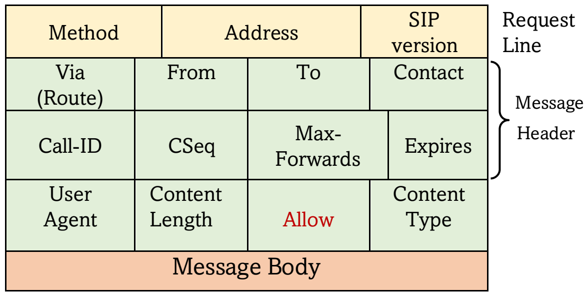

Fig. 1. SIP Request message showing the "Allow" field in the Message Header (red colored).

Fig. 2 depicts the message flow between two SIP terminals, and their nearest VoIP-servers (Asterisk and Centile, allowing SIP). The servers are two real industrial implementations between the two SIP-phones (caller and callee). The messages between these two servers (red colored in Fig. 2) result from the messages sent/received from/to both phones. The "Allow" field that includes the type "MRTP" is sent within the SIP Invite (first message in Fig. 2).

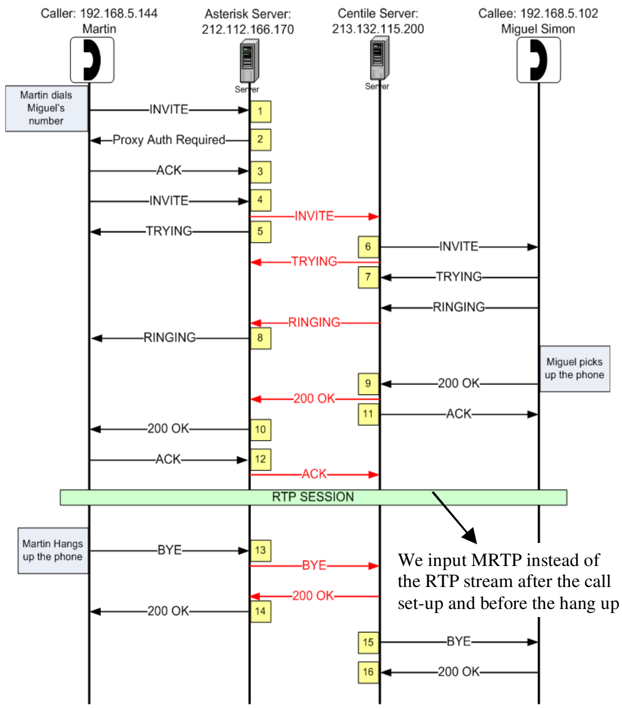

Fig. 2. The MRTP within the SIP Call message flow. It falls exactly in place of the RTP messages (real test example) i.e., the servers do not realize any difference in call set-up/hang-up.

If the MRTP was enabled on the other end (callee), a 200OK message (messages 9 in Fig. 2, which is the reply message from the callee agreeing to the communication)

 https://doi.org/10.1051/e3sconf/202235101075

shall also include "MRTP" in its "Allow" field. Then, the clients would know that they can communicate with MRTP enabled. If the Callee sends a 200OK without including the "MRTP" type, all nodes switch to RTP.

2.2 Media Protocols

Many standardized protocols can carry media including DVB (Digital Video Broadcasting), H.261, H.263, RTCP (RTP Control protocol), and RTP (Real-Time Transport). We choose RTP and RTCP [12] since they are the most suitable to use with SIP for QoS purposes [13]. In our setup, when medical data is to be transmitted, the solution uses the MRTP instead of RTP via informing the receiver through the "Allow" field in the SIP Request message Header (Fig. 1) as discussed above.

3 The MRTP Design

The MRTP is designed to be incapsulated in UDP packets (Fig. 3), and it has its own mechanism of checking packet loss via keeping MRTP-packet-counters and sending acknowledgment signals (MRTP-ACK) to the sender as part of the MRTP data.

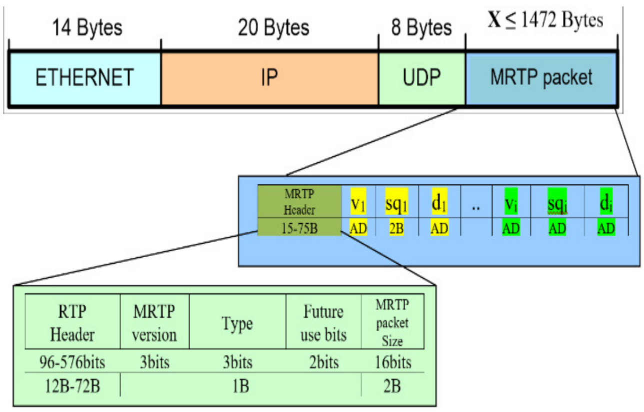

Fig. 3. MRTP over SIP data over UDP datagram. MRTP is compatible with the RTP protocol. AD stands for 'Application Dependent' size (bytes) for the indicated filed. v_i stands for volume i of the i^{th} data (of the i^{th} application). d_i stands for the data of the i^{th} application. sq_i stands for the sequence number for d_i. If the Type is '000' the MRTP packet is a handshake packet.

Every packet received by the phones, is made up of three main headers: Ethernet, IP and UDP. The UDP header is 8-Bytes and specifies the length of the payload. The MRTP packet contains the MRTP header and data. The header is backward compatible with RTP, where its first part contains the exact bits of the RTP header, and the rest includes the MRTP-version, type, and MRTP packet-size. The MRTP-data encapsulates the audio and medical information in sequenced-chunks, where each chunk includes 3 fields: volume (v_i) that informs of the size (bytes) of the i^{th} application data (e.g., an ECG signal could take a few bytes while an audio-message can be larger), the actual i^{th} application data (d_i), and the sequence number for d_i (sq_i). This sequence number is used in the piggy-backed acknowledgment (MRTP-ACK) for the sender to know if the data d_i was received or lost (i.e., should be retransmitted). This design allows sending data of multiple applications in one packet (one overhead) while keeping track of received/lost dataset (d_i) via tracking its sq_i. It allows to shift between MRTP and RTP based on whether the client has the MRTP solution installed or not.

4 Experimental Results

The experiments are set between two clients (a patient and MD) connected via two SIP phones (each on a separate network and server) that establish media sessions using 'RTP and TCP together' and MRTP (Fig. 2). Experiments are divided into two sets: only-audio experiments where we test the efficiency of MRTP compared to RTP when the phone is used for telehealth voice-consultation only, and 'audio with medical data' (data of 12 Human-Heart ECG leads) when the session has a patient-doctor consultation with medical data transmitted to the MD as well.

The delay is measured as back-to-back latency of two consequent packets. For instance, in the second set of experiments, the delay is measured between consequent RTP packets (carrying audio) and TCP packets (carrying medical data in text format) that are sent concurrently but on two different streams.

4.1 Only Audio Experiments

Since RTP is designed and used to carry real-time media, it is important to compare its performance in carrying audio to that of MRTP when carrying only audio. The tests are run for LAN-setup and international calls. Both showed better performance of MRTP compared to RTP. However, the LAN results are presented since LAN tests minimize external factors that may affect the characterization of bandwidth and other traffic issues i.e., to have the least external effects on the tests. To run the experiments, the average call length of 2.2 minutes is used for the real-time session (Fig. 2). We ran 1000 experiments (calls) in every set (each call lasting 2.2 minutes).

The results show that, even with only audio transmission, the MRTP average delay is 9.2% less than that of RTP for the same audio packets (Fig. 4).

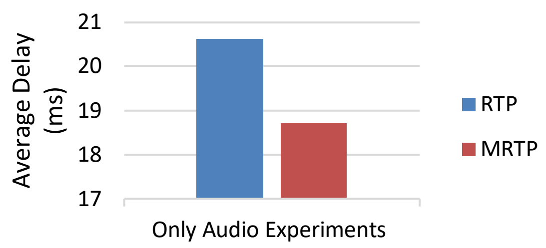

Fig. 4. Average delays for Only-Audio Experiments using RTP and MRTP over a LAN. The average delays are calculated for 1000 experiments for each set.

4.2 Concurrent Audio and Medical Data Experiments

In this set of experiments, audio and medical datasets are transmitted in parallel between the patient and MD. One set is done by sending the audio call over RTP and the 12 ECG-leads over 12 different TCP connections. Another set is done via transmitting the same audio and medical dataset in one MRTP packet (Fig. 3). Since the experiments check the performance of MRTP in comparison with combining two protocols (RTP and TCP) to send same datasets, this scenario is stretched to international calls to check for longer VoIP network paths (international hops).

The results (Fig. 5) show that the average delay of MRTP is around 74.1% less than that when concurrently utilizing RTP for audio-streams and TCP for medical-data transmissions (having 12 TCP connections; one connection per ECG lead).

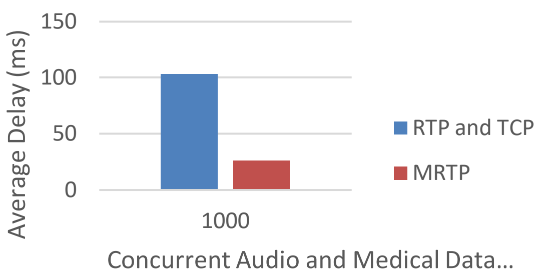

Fig. 5. Average delays for 'Audio over RTP with Medical Data over TCP' in comparison to 'Audio and Medical Data' over MRTP. The averages are for 1000 experiments for each set.

5 Conclusion

This paper presents a novel real-time protocol, MRTP, for enhancing online healthcare meetings and monitoring, which is a growing need especially after the COVID-19 pandemic. The MRTP minimizes real-time session delays for both, LAN-based telehealth-sessions and international ones. We utilize the SIP messaging protocol for call setup and hang-up due to its wide market deployment and its few call-setup messages. SIP low number of messages matches with our goal to minimize overhead i.e., decreasing delays. To achieve the goals, we modified the SIP Request message header by adding the "MRTP" type as a new option to its "Allow" field. This enables servers and clients to differentiate between telehealth sessions that utilize MRTP and normal calls with RTP. They can also check whether a packet carries medical-data or not. Our solution leverages the performance of telehealth applications that need real-time distance monitoring via decreasing average delays of the communicated streams/data. This solution paves the way for novel Internet Telephony scenarios with audio and text-data. The paper describes the changes in the SIP Request message Header (Fig. 1), fitting MRTP in place of RTP (Fig. 2), and the MRTP design and configuration (Fig. 3).

We compare the MRTP deployment results with those of RTP as well as with those of the concurrent use of RTP and TCP. The results show that for only-audio packets, the MRTP delay is 9.2% less than that when using RTP for the same audio packets (on the same LAN). Moreover, the MRTP design and implementation improve the performance by decreasing the average real-time session delay by 74.1% compared to the concurrent use of RTP (audio transmission) and TCP (12 ECG-lead transmission) over an international network.

This significant decrease in delay can help decrease traffic-loads and packet loss over telehealth networks, which would lead to better QoS and user (MD and patient) satisfaction.

References

1. Kastrenakes, J., Zoom saw a huge increase in subscribers — and revenue — thanks to the pandemic: Everyone's using Zoom, and it shows, The Verge, Tech (2020), https://www.theverge.com/2020/6/2/21277006/zoom-q1-2021-earnings-coronavirus-pandemic-work-from-home, last accessed 2021/05/19
2. Marin, A., Telemedicine takes center stage in the era of COVID-19, Technology Feature, Science Magazine (2020), https://www.sciencemag.org/features/2020/11/telemedicine-takes-center-stage-era-covid-19, last accessed 2021/07/24
3. Jiménez-Rodríguez, D., García, A. S., Robles, J. M., Salvador, M. d. M. R., Ronda, F. J. M., and Arrogante O., Increase in Video Consultations During the COVID-19 Pandemic: Healthcare Professionals' Perceptions about Their Implementation and Adequate Management, Int. J. of Env. Res. and Pub. Health **17**(**14**) (2020), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7400154/, last accessed 2021/07/24
4. Bestsennyy, O., Gilbert, G., Harris, A., Rost, J., Telehealth: A quarter-trillion-dollar post COVID-19 reality?, McKinsey & Company (2021), https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/telehealth-a-quarter-trillion-dollar-post-covid-19-reality#, last accessed 2021/07/24
5. Facts and Factors. Global Telehealth Market Size 2021 Rise at 26.5% CAGR, Will Reach USD 475.50 Billion By 2026: Facts & Factors, Intrado Globe Newswire (2021), https://www.globenewswire.com/news-release/2021/06/29/2254645/0/en/Global-Telehealth-Market-Size-2021-Rise-at-26-5-CAGR-Will-Reach-USD-475-50-Billion-By-2026-Facts-Factors.html, last accessed 2021/07/24
6. Keach, S., LOG OFF Zoom DOWN in mysterious outage as video calling app stops working around the world, The Sun, Tech (2020), https://www.thesun.co.uk/tech/11232380/zoom-down-not-working-offline-cant-connect/, last accessed 2021/05/19

7. Clark, D., *A taxonomy of internet telephony applications*, Proceedings of the 25th Telecommunications Policy Research Conference. Washington, DC, USA. (1997)
8. Advanced VoIP: VoIP QoS Version 1.0, Report, 6 (2006), https://docplayer.net/1498951-Voip-qos-version-1-0-september-4-2006-advancedvoip-com-sales-advancedvoip-com-support-advancedvoip-com-phone-1-213-341-1431.html, last accessed 021/05/19
9. Singh, K., and Schulzrinne, H., *Peer-to-peer internet telephony using SIP*, Proceedings of the international workshop on Network and operating systems support for digital audio and video, NOSSDAV '05, Skamania, Washington, USA. (2005)
10. Curcio, I. D., and Lundan, M., *SIP call setup delay in 3G networks*, Proceedings of the Seventh International Symposium on Computers and Communications pp. 835-840, Taormina, Italy, ISCC'02 (2002)
11. Pack, S., and Lee, H., Call Setup Latency Analysis in SIP-Based Voice over WLANs. IEEE Com. Letters, **12**(**2**), pp. 103-105 (2008)
12. Zhang, G., Towards Secure SIP Signaling Service for VoIP Applications, (Licentiate thesis), Royal Institute of Technology (KTH), Stockholm, Sweden. (2009)
13. Verma, D. C., Zhang, H., and Ferrari, D., Delay jitter control for real-time communication in a packet switching network, (Technical Report TR-91-007), Uni. of Cal., Berkeley, CA, USA. (1991)