

An Adaptive Videoconferencing Framework for Collaborative Telemedicine

Ronnie Muthada Pottayya
University UBFC
FEMTO-ST UMR 6174 CNRS
16 Route de Gray
25000 Besancon, France
Email: rmuthada@femto-st.fr

Jean-Christophe Lapayre
University UBFC
FEMTO-ST UMR 6174 CNRS
16 Route de Gray
25000 Besancon, France
Email: jc.lapayre@femto-st.fr

Eric Garcia
Ido-In (Maincare Solutions Group)
4 Rue du Pr. Paul Milleret
25000 Besancon, France
Email: eric.garcia@ido-in.com

Abstract—In the medical area, most of medical facilities (hospitals, clinics, ...) use distributed applications, such as telediagnosis for example. As information security is mandatory, these applications must be able to cross the security protocols (secured gateways like proxies, firewalls ...). User Datagram Protocol (UDP), which is classically recommended for videoconferencing applications, does not cross firewalls or proxies unless explicitly configured fixed ports are declared. These fixed ports are considered as a security breach. In this paper, we propose a new and novel platform called VAGABOND (Video Adaptation framework, crossing security GATeways, Based ON transcoDing) which works, in a very efficient and original way; on TCP (Transmission Control Protocol). VAGABOND is composed of Adaptation Proxies (APs), which have been designed to take into consideration medical experts videoconferencing preferences, device heterogeneities, and network dynamic bandwidth variations. VAGABOND is able to adapt itself at the user and network levels. The cumulative binomial probability law and the Bayesian inference on a binomial proportion are used to trigger user profile adaptations. Probabilities are calculated based on the number of retained video packets on all those received in a given lapse of time. A Bayesian inference is used to calculate a posterior probability by taking into account a prior and a likelihood. With the Bayesian probability interpretation, Baye's theorem expresses how a subjective degree of belief should rationally change to account for evidence. In fact, we aim at being more tolerant to severe network bandwidth variations. With a finer precision and following these probability laws, user profile adaptation is only triggered when severe network congestions arise. The framework uses TCP for the delivery of videoconferencing data. The advantage of this protocol is its ability to spontaneously cross firewalls and proxies (use of TCP ports 80 or 443). However, as TCP is a reliable transport protocol, we needed to design and to employ new intelligent adaptation strategies together with data transmission in order to cope with latency issues and sockets timeout.

Key words: Distributed adaptation, Proxies, Video on TCP, Binomial law, Bayesian inference, Telemedicine, and Videoconferencing.

I. INTRODUCTION

Videoconferencing allows individuals to interact and communicate visually without any encounter. In the medical field, by using this technology, doctors can help patients who are unable to physically visit hospitals, especially in rural areas where medical specialists are not always available. The importance of developing videoconference frameworks for health

applications have been mentioned in many previous works such as in the papers [1], [2], [3]. For example, in [3], the authors mentioned that the challenge is to find solutions that give room for expressiveness but can also be made available for a broad audience. For them, videoconferencing can be very useful for advice-giving over distance, specifically for issues that require a higher level of communicative expressiveness than the telephone can offer.

However, the deployment of videoconferencing solutions in medical centres is a challenging matter. Indeed, these types of infrastructures hold and manipulate everyday sensible patient medical records and hence one of the security policies is such that User Datagram Protocol (UDP) packets are usually all blocked by firewalls. The reason behind is that UDP-based systems are exposed to many security threats. One of them is the UDP-based flooding which is a common form of Denial of Services (DoS) attacks [4].

It is well known that almost all classical videoconferencing systems use the UDP protocol at the transport layer and the Real Time Protocol (RTP) at the session layer. Unfortunately, these protocols will not pass through firewalls and proxies set up in medical centres unless explicit ports allocations are declared on firewalls but UDP ports are considered as a security breach. If ever in a particular medical centre, ports allocations have been made on firewalls, UDP packets need to bypass proxies and will be routed through sub networks. As such, UDP-based videoconferencing systems actually present in medical centres operate separately on VLANs¹ where medical records are not found. In the case of integrated telemedicine systems, where we manipulate videoconferencing and medical data, the use of UDP protocol is not possible. We need a completely secured network where we can encrypt and encapsulate all medical data such as imagery and videoconferencing data.

In order to overcome all these challenges, we have designed and implemented an original and efficient system being able to perform adaptation at the user and network levels. This system is called VAGABOND (a Video Adaptation framework, crossing security GATeways, Based ON transcoDing). Based on our literature review, no works have proposed solutions to tackle this issue. We proposed the state of the art as a solution to solve this problem by using VAGABOND which can perform these adaptations at the user and network levels. Context awareness

¹Virtual Local Area Networks

was also introduced based on the videoconferencing needs and resource availability of the underlying transmission medium.

Our first performance tests have enabled us to validate the system. A phase of integration in the Covotem^{TM2} telemedicine platform, which is actually distributed in French hospitals, will begin in the forthcoming months. The aim of this article is to present this videoconferencing system, which has been tested in a hospital context. This system is able to adapt itself following medical experts videoconferencing preferences, user device and network access heterogeneities, dynamic bandwidth variations, and random network impairments.

The rest of this article is organized as follows: part two presents a literature review, part three presents the VAGABOND framework together with its evaluation. Finally, part four concludes this paper.

II. LITERATURE REVIEW

Video telephony had little success during the last decade due to its stringent bandwidth and delay requirements. For example, in live streaming video, buffering delay is often tolerable whereas in videoconferencing, the user Quality-Of-Experience (QoE) degrades significantly if the one-way end-to-end video delay goes over 350 milliseconds [5]. Moreover, videoconferencing solutions have to cope with user device and network access heterogeneities, dynamic bandwidth variations, and random network impairments, such as packet losses and delays.

A. Telemedicine

JB Aupet *et al.* [6], [7] defined telemedicine as the use of telecommunication technologies to provide medical information with services. This process uses electronic, visual and audio communications to support practitioners at remote sites with diagnosis and consultation procedures, such as clinical examinations and medical image transfers. In telemedicine, formalization of information exchanged between medical experts is very important because it facilitates a transparent and traceable understanding of the telemedical processes [8]. In a recent paper [9], the authors highlighted some key aspects of telemedicine. They stated that telemedicine is an umbrella term that encompasses any medical activity involving an element of distance. This definition is backed by [8]. They also mentioned that mobile telemedicine applications need to overcome several problems associated with users' mobility. B. Kamsu-Foguem *et al.* [10] distinguished sub-medical activities related to telemedicine. The authors defined telemedicine as being a remote practice, which utilizes advanced telecommunications and information technologies for the delivery of healthcare and the exchange of health information across distances. The sub-medical activities for them are: teleconsultation, teleexpertise, telemonitoring and teleassistance. We will stick to these definitions in our present work.

B. Off-the-shelf videoconferencing solutions

Telemedicine is a particular field where off-the-shelf videoconferencing solutions are not adapted to this context. For

instance, the authors in [11] reported that the user Quality of Experience (QoE) ranged between 54 and 80 percent when Apple FaceTime was used in a teledermatology context. In [12], [13], the authors measured the responsiveness of Skype video calls to bandwidth variations and argued that Skype's response time to this kind of phenomenon is too long. Furthermore, Skype remains an unsecured videoconferencing solution, predominately in the health sector where security is a key aspect [14].

C. Telemedicine videoconferencing solutions literature review

In the literature, few studies have been conducted especially addressing the aforementioned issues and challenges of video consultation sessions in low bandwidth networks, being able to cross security gateways easily, and taking into account users preferences.

An early study was conducted by G. K. Chan *et al.* [15] on a secure multi-access, cross-platform telemedicine application, MEETING ROOM. It is based on Adobe Flash technology. It is a hybrid one, comprising of a web portal for patient medical data and a videoconferencing application which adopts a web-based approach in general to provide the flexibility to run on cross-platform devices. However, in MEETING ROOM, the Real-Time Messaging Protocol (RTMP) is used which implies the use of a fixed port 1935. Besides, adaptation mechanisms are neither present at the user level nor at the network level in MEETING ROOM. P. Rodríguez *et al.* [16] proposed a cross-device videoconferencing solution for wireless networks based on Quality of Service (QoS) monitoring. Their solution is named VaaS and the technologies employed are the same as MEETING ROOM. The difference compared to MEETING ROOM is that an algorithm has been designed and it decides how to adapt the current video resolution at any given time considering the input obtained from the network monitoring system found in VaaS. Both systems network topology is a client-server one and use TCP as the transport protocol.

B. Parmanto *et al.* [17] developed a versatile system for tele-rehabilitation (VISYTER). This system supports audio and video data. However, UDP is used for the delivery of videoconferencing data and the system is designed to be used in broadband networks and would not support narrowband communication when the bandwidth drops below 128 kbps. Hoeng Helen *et al.* [18] concluded that an improved technology and infrastructure are needed to better meet the clinical requirement of telehealth. Their study was based using an off-the-shelf teleconferencing solution and the results show that narrow bandwidths lesser than 384 kbps severely affect the reliability of assessment accuracy.

An interesting telemedicine system architecture is given in [19]. The authors made an overview of recent end-to-end wireless medical video telemedicine systems using the 3G network. They made a clear classification of research in mobile health (m-health) and depicted some Region of Interest (ROI) based systems. Nevertheless, the system does not take into account user preferences and besides the system is only designed to transmit real-time medical ultrasound (or trauma) video.

As Wei-Li Lui *et al.* [20] mentioned in their study on internet-based videoconferencing coder/decoders and tools for

²<https://www.ido-in.com/solutions/telemedecine-tele-surveillance/telemedecine-et-tele-surveillance,282,219.html>

telemedicine, some applications may be completely inadequate for certain types of telemedicine. For example, in teledermatology, higher video resolution may be needed to show a patient’s skin more clearly. In this particular case, an audio stream may be useless. Hence, more bandwidth can be allocated to the video stream. Each of the work discussed in this paper attempts to address one or a few issues whereby leaving the others compromised. Solutions that are really pulled by real clinical requirements and thus embodying the medical domain of knowledge into the overall framework and solution are more convenient. It calls for synergy from efforts of both medical and technological practitioners to have systems like that. VAGABOND has been designed for medical practitioners and aims at bridging the gaps between off-the-shelf solutions and the requirements of telemedicine.

III. CONTRIBUTION

A. Proposition

Adaptation at the user’s level: VAGABOND has been designed to take into consideration health professionals videoconferencing needs and we set up a non-exhaustive database from them. In the application, a user defines his profile and according to it, the system will automatically apply the corresponding videoconferencing preferences whenever a network congestion is detected by the network adaptation module (see section III-B). These preferences are found in a central database on the AS and are downloaded during the setup process of a videoconferencing session.

An example is in teledermatology context where diagnostics of tumoral dermatosis require high video resolution and precision. Indeed, a snapshot of the analyzed part of a patient’s skin can be taken at any time for image thresholding and contours extraction. Consequently, the video does not need to have a high framerate. His own video is not required in this kind of teleconsultation but voice communication is needed for him to give instructions to the patient. These characteristics are found in the database and can be summarized as follows:

- Teledermatology → Patient(High video resolution, low framerate, audio), Expert(audio only)

Sometimes, face-to-face meeting with the medical experts are necessary. As it is the case in neurology. A neurologist needs to ask a patient to make specific gestures for clinical and motor skills diagnosis. As such, the video resolution of the patient needs to be fluid. In this context, only a limited view of the expert can be useful, meaning that a low video quality is acceptable. These characteristics can be modeled in the database as follows:

- Neurology → Patient(low video resolution, high framerate, audio), Expert(low video resolution, high framerate, audio)

Adaptation at the network’s level: There are three main architectures, which are employed for videoconferencing solutions: client-server, peer-to-peer, and a mix of the two. VAGABOND’s network topology is purely a client-server one. It has a main server, the Adaptation Server (AS), which is in charge of registering sessions and allocating them to clients. The AS is the entry point for all clients. Clients do not have

access directly to sessions. In doing so, we alleviate clients of the burden of keeping traces of all sessions locally on their sides. When a session appears or disappears, it will notify this event only to the AS. Thus, sessions do not have to inform clients about their availability. In VAGABOND, a session is an adaptation proxy (AP). It is the server to which different clients will exchange their video and voice data. AS and AP are always found on public IP addresses. Secured communications are always set up and all exchanges between clients, AS, and APs are encrypted using AES³. Figure 1 shows how sessions

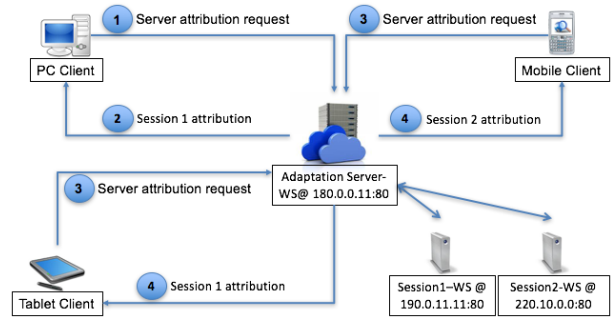


Figure 1: Sessions attributions

are allocated to clients. When a client enters a meeting, it first sends a session allocation request to the AS (steps 1 and 3 from figure 1). The request encompasses the user’s device information, an identifier of the meeting room the user is willing to enter, and a network bandwidth estimation at the time of the request. Bandwidth estimation is calculated by a client before an allocation request using the AS. An iPerf⁴ server is deployed on the AS. iPerf is a tool for active measurements of the maximum achievable bandwidth on IP networks. It supports tuning of various parameters related to timing, buffers and protocols (TCP, UDP, SCTP with IPv4 and IPv6). For each test, it reports the bandwidth, loss, and other parameters such as the Maximum Transmission Unit (MTU) size and observed read sizes.

Upon receiving an allocation request, the AS will search for an AP server following the user’s bandwidth capacity. If the bandwidth estimation of the client is relatively low, the AS will allocate to the latter an AP with a high bandwidth capacity in order to reduce latencies induced by the client’s available bandwidth. If the meeting room does not exist, the AS will create one and register the requesting user into it. The AS will respond to the client’s request by sending to it, the best session (AP), which corresponds to its properties (steps 2 and 4 from figure 1). In this response, will be found information like the contact information of the session attributed (AP), and information about other participants present in the requested meeting room. The connection between clients and the AS is a persistent (stateful) one.

Upon receiving a response from the AS, a client analyzes it to know whether it is alone or not in the meeting room. If it is alone, no sharing of data is done. Else, it will share its video and audio streams to its corresponding AP. Clients receive notification messages whenever a client enters or leaves

³Advanced Encryption Standard

⁴<https://iperf.fr>

a meeting room from the AS. Clients will start sharing their

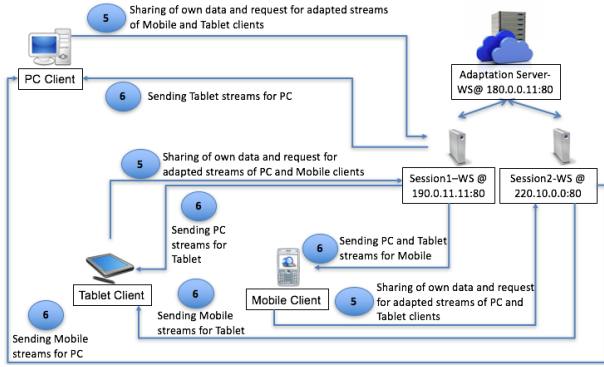


Figure 2: Data sharing between participants

videoconferencing data and will ask their respective session for other participants' videoconferencing data. A client continuously shares its own data with its corresponding session (step 5 from figure 2) and the latter sends these adapted data to other participants of the meeting room (step 6 from figure 2). Indeed, an AP will adapt video data according to the client to which these data are intended. Video transcoding techniques are used. In figure 1, the "Mobile Client" was allocated with "Session 2" and is sharing its data to it. Sessions do not communicate with each other. "Session 1" is sending the "PC Client" and the "Tablet Client" adapted streams to the "Mobile Client". This is explained in the figure 3 where intermediates steps happened between step 5 and step 6.

When a client (let's call it, the requester) requests streams of other clients (step 5), which are not found on the same session (AP) as itself, its attached session will forward this request to the AS (step 5.1). Upon receiving this request from one of its session, the AS will search for the sessions where the clients are found and communicate these information to the requesting session (step 5.2). On receiving these information, the requester will initiate other connections with the sessions where the missing clients are found (step 5.3) and will request the sharing of data (step 5.4). The requester will continue to share its own data to its allocated session (in our figure, session 2) but will receive data of other participants from other sessions.

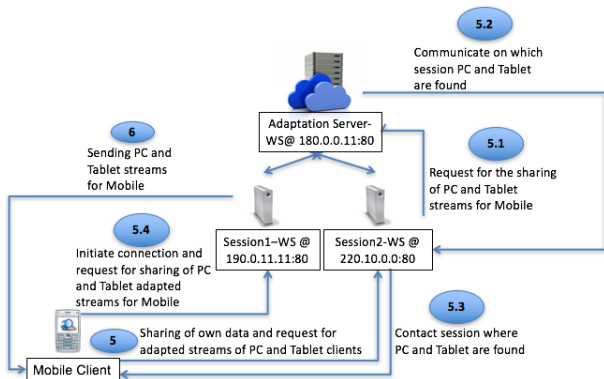


Figure 3: Automatic client's failover onto another session

B. User adaptation triggering based on video packets transmission

The network adaptation module on the client side is responsible for providing feedback information concerning the total number of frames received, those dropped, and those retained. It calculates separately these information for video and audio packets. Dropped audio packets are not discussed in this paper as audio packets are generally small packets and are not the major cause of bottleneck effects. On the contrary, video packets are the major cause of bottlenecks due to their sizes. Furthermore, dropping video frames can cause a videoconferencing session to have a lower framerate but will still be usable at a certain tolerance threshold. For instance, dropping 50% of received audio frames may result in an unusable videoconferencing session whereas for video frames, it can still be usable. A graceful degradation of the video will be noticed. The information about video packets reception is taken into account every 10 seconds and a probability of triggering an adaptation is calculated for the next 3 seconds. The period of 10 seconds was chosen as it enables us to have enough samples and the period of 3 seconds to be quickly reactive. This time is actually variable and not fixed. Further tests with users will enable us to refine these values. As shown in figure 4, a client receives packets from the network and whenever the difference between the time stamp in a packet and the client's time stamp is larger than 700 milliseconds, this packet is dropped and another one is taken into consideration. The interval time step of 700 milliseconds has been chosen as, according to performance tests conducted by Jansen *et al.* [5] over the Internet, it is the delay suitable for a high quality videoconferencing system. The optimal tolerable delay for VAGABOND is left for us as a future work as further tests with users (users situated behind a proxy and/or firewall) will enable us to refine this value.

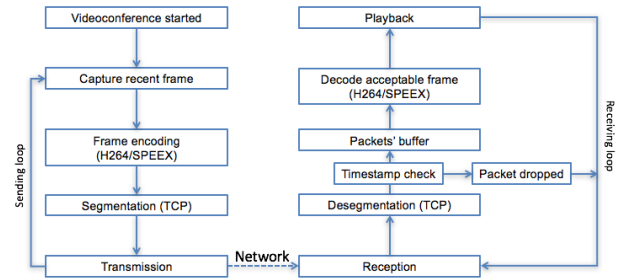


Figure 4: Sending and receiving videoconferencing frames

User profile adaptation is triggered firstly using the binomial probability law:

$$Pr(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad (1)$$

Suppose that an expert client received 80 video packets in a cycle of 10 seconds. The information monitor detects that only 25 packets are retained out of the 80 packets. The rate of success is 0.3125 (0.6875 failure). We are in the presence of a Bernoulli scheme and the random variable X taking for values the number of successes follows the binomial law with parameters $n = 80$ and $p = 0.3125$ noted $X \sim B(n, p)$. The

probability of getting less than 24 successes ($\frac{80}{10} \times 3seconds$) in the 80 trials is given by the cumulative distribution function:

$$\begin{aligned} F(24; 80, 0.3125) &= Pr(X \leq 24) \\ &= \sum_{i=0}^{24} \binom{80}{i} 0.3125^i (1 - 0.3125)^{80-i} \\ &= 0.4580 \end{aligned} \quad (2)$$

Hence, we have here a probability of 0.4580 of triggering a user profile adaptation. The expert will experience a graceful degradation of the video but it will still be usable. Therefore, an adaptation will not occur.

Suppose now, that only 31 frames out of 104 frames are retained on the next 10 seconds cycle. The rate of acceptance is 0.2981 (0.7019 failure). Consequently, calculating a cumulative binomial probability of triggering an adaptation for the next 3 seconds gives us:

$$\begin{aligned} F(31; 104, 0.2981) &= Pr(X \leq 31) \\ &= \sum_{i=0}^{31} \binom{104}{i} 0.2981^i (1 - 0.2981)^{104-i} \\ &= 0.5481 \end{aligned} \quad (3)$$

When the cumulative binomial distribution function gives a probability of more than 0.5 (meaning that an adaptation is required), Bayesian inference of the binomial proportion is used [21], [22]. The idea is to calculate a posterior probability based on a prior and a likelihood. The prior distribution that incorporates our subjective beliefs is based on the parameter of interest which is the number of retained packets in our case. The likelihood is the actual proportion of retained packets over all received packets (Bernoulli likelihood function). The basis of all Bayesian statistics is Baye's theorem :

$$posterior \propto prior \times likelihood \quad (4)$$

In our case, the likelihood is binomial. If the prior and the posterior distribution are in the same family, the prior and posterior are called conjugate distributions. We use the beta distribution which is a conjugate prior because the posterior is also a beta distribution. Furthermore, the beta distribution is defined on the interval $[0, 1]$ and possesses two shape parameters known as α and β , which give it significant flexibility. This flexibility provides us with choices in how we model our beliefs. We say that the beta distribution is the conjugate family for the binomial likelihood. In this instance, the probability density function of the beta distribution is given by the following:

$$P(\theta|\alpha, \beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)} \quad (5)$$

Where the term in the denominator, $B(\alpha, \beta)$ is present to act as a normalizing constant and gives us the prior belief distribution. Using the beta distribution, we only have to specify two parameters, α and β . These two parameters neatly correspond to the mean and variance of the beta distribution. The mean of previously retained video packets is given by:

$$\bar{\pi}_{Prior} = \frac{\alpha}{\beta} \quad (6)$$

And the prior variance given by:

$$\widehat{Var}(PrV)_{Posterior} = \sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}} \quad (7)$$

Where:

- α is the number of prior retained video packets,
- β is the number of prior dropped video packets

Since, our goal is to calculate a posterior probability by taking into account a prior and a likelihood, the posterior beta distribution is simply:

$$Beta(\zeta + \alpha, N - \zeta + \beta). \quad (8)$$

Where:

- N is the number of actual received video packets,
- ζ is the number of actual retained video packets

The posterior mean is therefore:

$$\bar{\pi}_{Posterior} = \frac{\zeta + \alpha}{N + \alpha + \beta} \quad (9)$$

And the posterior variance:

$$\widehat{Var}(PoV)_{Posterior} = \sqrt{\frac{(\zeta + \alpha)(N - \zeta + \beta)}{(N + \alpha + \beta)^2(N + \alpha + \beta + 1)}} \quad (10)$$

As an example, we will take the cumulative binomial probability obtained from equation 3 at time t . Relying only on this value, we would trigger a user profile adaptation. The mean, at time $t-1$, was calculated to be 0.7901 for 64 packets retained out of 81 with a standard deviation of 0.045. This gave us the prior belief distribution of $Beta(\omega|64, 17)$. At time t , we observed 31 retained packets out of 104. Our posterior beta distribution is:

$$\begin{aligned} Beta(\omega|\theta + \alpha, N - \theta + \beta) &= Beta(\omega|31 + 64, 104 - 31 + 17) \\ &= Beta(\omega|95, 90) \end{aligned} \quad (11)$$

At this stage, we can compute the mean and standard deviation of the posterior in order to produce probability estimates on the reception of the next set of video packets within 10 seconds. In particular, the value of $\bar{\pi}_{Posterior}$ is 0.5135 using equation 9 while the standard deviation $\widehat{Var}(PoV)_{Posterior}$ is 0.0366 using equation 10. A mean of $\omega = 0.5135$ states that approximately 51.35% of the packets received will be retained in the next 10 seconds while the standard deviation of 0.0366 means that we are uncertain about this 51.35% by 3.66%.

An automatic user profile adaptation is applied when more than 50% of the received video packets are dropped. This percentage is actually variable and not fixed. Further tests with users will enable us to refine this value. When an automatic user profile adaptation is triggered, a client will send its status to its AP which in turn will broadcast this notification to all the clients connected in the same meeting room as the one experiencing a network congestion. For example, in a teledermatology context, experts in the meeting room will only have their audio activated and patients will have their audio and video activated with a higher resolution but with a lower frame rate than the original one.

C. Experimental results

We compared the results of using the binomial distribution law against a frequentist approach when applying a user profile adaptation. For instance, in equation 2, a frequentist approach will trigger a user profile adaptation with a probability of 0.6875 whereas using a binomial distribution approach, we have a forecast of the situation over 3 seconds and a probability of 0.4580. We carried out several experiments (with the user profile adaptation deactivated) using unstable networks (3G networks) and limited networks (using NetLimiter⁵). Figure 5 shows the results of an experiment we carried out on an unstable 3G network. In this figure, it can be noticed that a user profile adaptation would have been triggered at time, $t = 10$, if only a frequentist probability was used. Results prove that a frequentist approach is not tolerant with dynamic bandwidth variations. On the curves shown in figure 5, a user profile adaptation would have been triggered at time, $t = 80$, using a cumulative binomial distribution probability (blue curve). However, the Bayesian inference of the binomial proportion predicted that 51.35% (see equation 11) of all received packets in the next 10 seconds will be retained. Hence, a user profile adaptation triggering did not occur. It was only triggered at time $t = 90$, with a cumulative binomial probability of 0.6367 and with a Bayesian inference of 29.64% (with $Beta_{Prior}(\omega|95, 90)$ and $Beta_{Posterior}(\omega|59, 140)$).

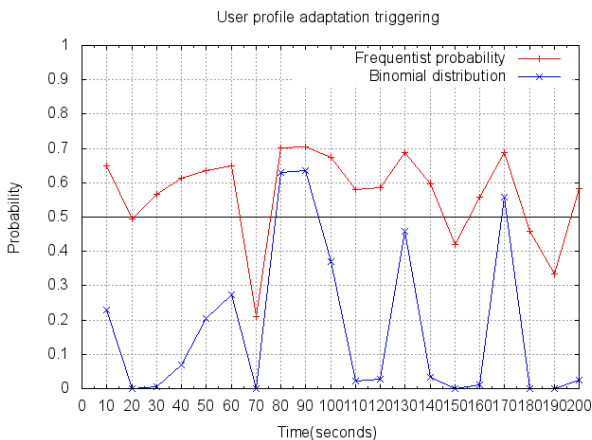


Figure 5: Comparison between frequentist probability and binomial distribution probability on a 3G network

Measurements were done on limited networks using NetLimiter without triggering any user profile adaptation. Our goal was to continue observing the evaluation of probabilities using the binomial law and the Bayesian inference on the binomial proportion. Figure 6 shows one of them. At time $t = 60$, the number of packets dropped began to increase. The frequentist probability was 0.5753 while the cumulative binomial probability was 0.0207. At time $t = 90$, the latter was 0.6499 with the Bayesian inference predicting only 30.43% of retained packets out of all received packets in the next 10 seconds. A user profile adaptation would have been triggered here.

Experiments on the number of packets dropped were

⁵<https://www.netlimiter.com>

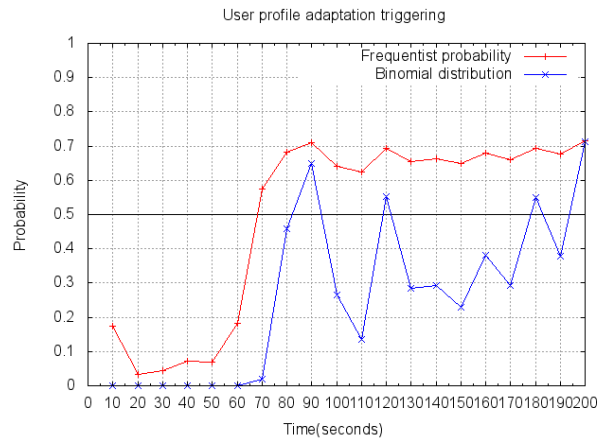


Figure 6: Comparison between frequentist probability and binomial distribution probability on a limited network

conducted when the adaptation module is activated and deactivated. Every 10 seconds, the total number of dropped packets were taken into account. A test bed using the Internet was set up with a process running on a working station and sending videoconferencing frames to two other processes running on another working station. One process had the adaptation module activated while for the other, it was deliberately deactivated. The idea was to compare the number of packets dropped with and without a user profile adaptation triggering (dermatologist profile used). Figure 7 shows the results of one of these experiments. At time, $t = 160$, an adaptation was needed. After $t = 160$, the red curve shows that the number of dropped packets is higher when no adaptation was performed. Almost every packet received was dropped. On the contrary, the blue curve shows that some packets were dropped but the videoconferencing session would still be usable when an adaptation was triggered.

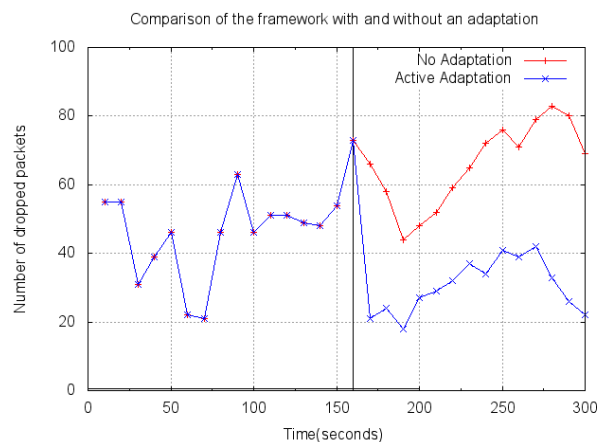


Figure 7: Comparison with and without a user profile adaptation

We also carried our experiments on end-to-end video packets delays of the system. The end-to-end video delay perceived by a user is the sum of delays incurred by real-time

video capturing, encoding, segmentation, transmission, and desegmentation. Let T_{CES} be the video capturing, encoding, and segmentation delay at the sender, T_N be the one-way transmission delay on the network between the sender and the receiver, T_S be VAGABOND's server processing time, and T_D be the desegmentation time. Thus, the one-way video delay is:

$$T = T_{CES} + T_N + T_S + T_D \quad (12)$$

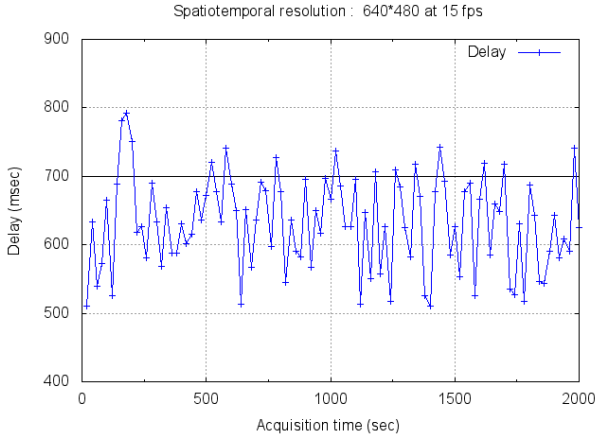


Figure 8: Delay v/s acquisition time graph for a spatiotemporal resolution of 640*480 @ 15 frames/second

In order to have an overview of the latency induced by using the framework with the TCP protocol, we performed realistic tests using the Internet between two working stations behind proxies and firewalls and located in two different hospitals with good bandwidths capacities. User profile adaptation was deliberately deactivated to continue having latencies. The difference between the captured frame time stamp and the desegmented packet time stamp is taken into account in our experiments. One of these experiments is shown in figure 8. Each test lasted 2000 seconds. From the values collected, a mean end-to-end latency time was calculated every 20 seconds. The mean delay for video frames was 646 milliseconds when tests were performed with a resolution of 640*480 at 15 frames per second with an average bandwidth usage of not more than 450 kbps.

Another experiment is given on figure 9. The mean delay for video frames was 655 milliseconds when tests were performed with a resolution of 1280*720 at 15 frames per second with an average bandwidth usage of not more than 650 kbps.

IV. CONCLUSION

The new and novel adaptive framework for multimedia applications in medical centres called VAGABOND is proposed in this article. This framework is an innovative approach in medical centres to handle videoconferencing data exchange in heterogeneous applications involving multiple hosts with different bandwidths and calculations capacities. VAGABOND's architecture consists of a main Adaptation Server (AS) and several Adaptation Proxies (APs) which can be added or removed. APs are chosen in respect to users' bandwidth estimation in order to cope with latency issues. Bandwidth-rich servers are used to relay videoconferencing data as soon

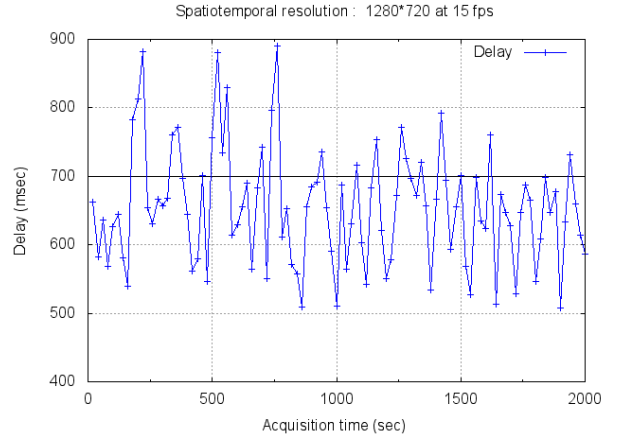


Figure 9: Delay v/s acquisition time graph for a spatiotemporal resolution of 1280*720 @ 15 frames/second

as they are received from bandwidth-poor clients. Furthermore, adaptation is realized and managed by VAGABOND at the user and network levels. The cumulative binomial probability law and the Bayesian inference on a binomial proportion are used to trigger user profile adaptations. Probabilities are calculated based on the number of retained video packets on all those received in a given lapse of time. A Bayesian inference is used to calculate a posterior probability by taking into account a prior and a likelihood. With the Bayesian probability interpretation, Baye's theorem expresses how a subjective degree of belief should rationally change to account for evidence. In fact, we aim at being more tolerant to severe network bandwidth variations. With a finer precision and following these probability laws, user profile adaptation is only triggered when severe network congestions arise.

This framework aims at helping health professionals in their medical diagnosis by applying automatic adaptation techniques. Focus is put mainly on patients' diagnosis by health professionals rather than trying to find the best configurations for a videoconferencing session. Moreover, a key aspect of providing a reliable and secured collaborative telemedicine videoconferencing solution for healthcare institutions, where networks are highly restrictive and UDP packets are generally blocked, has been met. The system as it runs today is complete and functional. Experiments proved that VAGABOND is tolerant to network bandwidth variations and that the system meets the acceptable video performance of 700 milliseconds end-to-end delay.

For our future work, we plan to make studies on the parameters used in VAGABOND in triggering a user profile adaptation following users' Quality of Experience (QoE). Such parameters are: the time at which the network adaptation module has to launch probability calculations, the next time step on which a probability will be forecasted, and the tolerable percentage of dropped packets before a videoconferencing session begins to be unusable. Conflicts between user profiles of medical experts will also be taken into consideration as with the actual system, experts with different user profiles cannot attend the same conference with patients. We believe that the use of ontologies in modeling medical experts profiles may be

the clue in resolving these conflicts.

REFERENCES

- [1] H.-H. Tsai, Y.-F. Tsai, H.-H. Wang, Y.-C. Chang, and H. H. Chu, "Videoconference program enhances social support, loneliness, and depressive status of elderly nursing home residents," *Aging and Mental Health*, vol. 14, no. 8, 2010, pp. 947–954.
- [2] H.-H. Tsai and Y.-F. Tsai, "Changes in depressive symptoms, social support, and loneliness over 1 year after a minimum 3-month videoconference program for older nursing home residents," *Journal of medical Internet research*, vol. 13, no. 4, 2011, p. e93.
- [3] M. Östlund, N. Dahlbäck, and G. I. Petersson, "3d visualization as a communicative aid in pharmaceutical advice-giving over distance," *Journal of medical Internet research*, vol. 13, no. 3, 2011, p. e50.
- [4] K. Verma, H. Hasbullah, and A. Kumar, "An efficient defense method against udp spoofed flooding traffic of denial of service (dos) attacks in vanet," in *Advance Computing Conference (IACC), 2013 IEEE 3rd International*. IEEE, 2013, pp. 550–555.
- [5] J. Jansen, P. Cesar, D. C. Bulterman, T. Stevens, I. Kegel, and J. Issing, "Enabling composition-based video-conferencing for the home," *IEEE Transactions on Multimedia*, vol. 13, no. 5, 2011, pp. 869–881.
- [6] J.-B. Aupet, R. Kassab, J.-C. Lapayre, and F. Marzani, "Adaptive image flow in collaborative medical telediagnosis environments," in *Computer Supported Cooperative Work in Design (CSCWD), 2011 15th International Conference on*. IEEE, 2011, pp. 685–689.
- [7] J.-B. Aupet, E. Garcia, H. Guyennet, J.-C. Lapayre, and D. Martins, "Security in medical telediagnosis," in *Multimedia Services in Intelligent Environments*. Springer, 2010, pp. 201–226.
- [8] R. Wootton, "Telemedicine," *British Medical Journal*, vol. 323, no. 7312, 2001, pp. 557–560.
- [9] N. N. Castellano, J. A. Gazquez, R. M. G. Salvador, A. Gracia-Escudero, M. Fernandez-Ros, and F. Manzano-Agugliaro, "Design of a real-time emergency telemedicine system for remote medical diagnosis," *Biosystems Engineering*, vol. 138, 2015, pp. 23–32.
- [10] B. Kamsu-Foguem, P. F. Tiako, L. P. Fotso, and C. Foguem, "Modeling for effective collaboration in telemedicine," *Telematics and Informatics*, vol. 32, no. 4, 2015, pp. 776–786.
- [11] R. Brandt and D. Hensley, "Teledermatology: The use of ubiquitous technology to redefine traditional medical instruction, collaboration, and consultation," *The Journal of clinical and aesthetic dermatology*, vol. 5, no. 11, 2012, p. 35.
- [12] L. De Cicco, S. Mascolo, and V. Palmisano, "Skype video congestion control: An experimental investigation," *Computer Networks*, vol. 55, no. 3, 2011, pp. 558–571.
- [13] X. Zhang, Y. Xu, H. Hu, Y. Liu, Z. Guo, and Y. Wang, "Modeling and analysis of skype video calls: Rate control and video quality," *IEEE Transactions on Multimedia*, vol. 15, no. 6, 2013, pp. 1446–1457.
- [14] E. De Keating-Hart, P.-Y. Milliez, and I. Auquit-Auckbur, "«skype® ambu»: la télé-médecine dans la chirurgie ambulatoire de la main," *Chirurgie de la Main*, vol. 33, no. 6, 2014, p. 458.
- [15] G. K. Chan, T. K. Kiong, and A. S. Narayanan, "Meeting room-a secure multi-access, cross-platform telemedicine application," in *Telecommunication Networks and Applications Conference (ATNAC), 2013 Australasian*. IEEE, 2013, pp. 201–206.
- [16] P. Rodríguez, A. Alonso, J. Salvachúa, E. Barra, and J. Cervino, "Adaptive cross-device videoconferencing solution for wireless networks based on qos monitoring," in *Computer and Information Technology (WCCIT), 2013 World Congress on*. IEEE, 2013, pp. 1–6.
- [17] B. Parmanto, A. Saptono, G. Pramana, W. Pulantara, R. M. Schein, M. R. Schmeler, M. P. McCue, and D. M. Brienza, "Visyter: Versatile and integrated system for telerehabilitation," *Telemedicine and e-Health*, vol. 16, no. 9, 2010, pp. 939–944.
- [18] H. Hoenig, L. Tate, S. Dumbleton, C. Montgomery, M. Morgan, L. R. Landerman, and K. Caves, "A quality assurance study on the accuracy of measuring physical function under current conditions for use of clinical video telehealth," *Archives of physical medicine and rehabilitation*, vol. 94, no. 5, 2013, pp. 998–1002.
- [19] A. Panayides, M. Pattichis, C. Pattichis, C. Schizas, A. Spanias, and E. Kyriacou, "An overview of recent end-to-end wireless medical video telemedicine systems using 3g," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. IEEE, 2010, pp. 1045–1048.
- [20] W.-L. Liu, K. Zhang, C. Locatis, and M. Ackerman, "Internet-based videoconferencing coder/decoders and tools for telemedicine," *TELEMEDICINE and e-HEALTH*, vol. 17, no. 5, 2011, pp. 358–362.
- [21] R. E. Mealey. *Obscure analytics*. [Online]. Available: <http://www.obscureanalytics.com/2012/07/04/to-the-basics-bayesian-inference-on-a-binomial-proportion/> (2016)
- [22] A. Etz. *Understanding bayes: Updating priors via the likelihood*. [Online]. Available: <https://alexanderetz.com/2015/07/25/understanding-bayes-updating-priors-via-the-likelihood/> (2015)