Evaluation of gatekeeper proxies for firewall traversal in secure videoconferencing systems

Prasad Calyam*, Gregg Trueb and Nathan Howes
Ohio Supercomputer Center,
The Ohio State University,
1224 Kinnear Road, Columbus, OH 43212, USA
Fax: 614-728-8110
E-mail: pcalyam@osc.edu
E-mail: gtrueb@osc.edu
E-mail: nhowes@osc.edu
*Corresponding author

Abstract: It is common today to have H.323 and SIP videoconferencing equipment deployed behind firewalls/NATs in campus and enterprise networks. A major challenge faced by network planners is to configure firewalls and gatekeeper proxies to allow voice-and-video traffic in-and-out of the internal-network’s ports while limiting malicious access of internal-network data by intruders through the same open ports. In this paper, we first describe the strategies used with gatekeeper proxies to solve the firewall traversal challenges in securing distributed videoconferencing systems. Next, we empirically evaluate the load-handling of gatekeeper proxies for firewall traversal under low, medium and high cross-traffic loads using subjective and objective measurements. Following this, we describe the signalling-and-multimedia flow architectures and identify caveats that arise due to heterogeneous adoption of these strategies. Lastly, based on our empirical results, caveats identification and vast operations experience, we list best-practices for deploying gatekeeper proxies in small-to-large scale secure videoconferencing systems.

Keywords: secure videoconferencing; firewall traversal; network middlebox; ITU-T H.460; gatekeeper proxy; video quality measurement.


Biographical notes: Prasad Calyam received his BS in Electrical and Electronics Engineering from Bangalore University, India and his MS and PhD in Electrical and Computer Engineering from The Ohio State University in 1999, 2002 and 2007, respectively. He is currently a Senior Systems Developer/Engineer at the Ohio Supercomputer Center, The Ohio State University. His current research interests include multimedia networking, cyber security, cyber infrastructure systems and network management.

Gregg Trueb received his BS in Electrical and Computer Engineering from The Ohio State University in 2009. While pursuing his BS, he was involved in several research projects at the Ohio Supercomputer Center. One of the projects that he contributed was a secure videoconferencing research project, whose salient findings are presented in this paper. Currently, he is employed at Northrop Grumman Corporation. His current research interests include multimedia networking and cyber-security.

Nathan Howes received his BS in Computer Science and Engineering from The Ohio State University in 2009. While pursuing his BS degree, he was involved in several research projects at the Ohio Supercomputer Center. One of the projects that he contributed was a secure videoconferencing research project, whose salient findings are presented in this paper. Currently, he is employed at Oracle Corporation. His current research interests include active/passive network measurements, multimedia networking and cyber-security.

1 Introduction

With increased access to broadband, internet videoconferencing has emerged as a viable medium for communication and entertainment. It is being used increasingly for applications such as remote meetings, distance learning, telemedicine and telemusic over the internet. The protocols involved for deploying such applications of videoconferencing have also been standardised to the point that they are being integrated into a
variety of communication devices ranging from mobile devices to high-definition video displays. The two major protocol standards used for videoconferencing are H.323 (ITU-T Recommendation H.323, 1999) and SIP (Handley et al., 1999) standards. The H.323 and SIP standards describe the codecs, devices and signalling schemes for supporting real-time voice and video communications such as videoconferencing over the internet.

A typical videoconferencing system consists of user end-point (EP) devices along with other devices such as:

a. firewall to control the incoming and outgoing traffic
b. multipoint control unit (MCU) a.k.a. soft-switch for supporting three or more simultaneous videoconference EPs
c. gatekeeper to provide services such as address information/translation and authorisation to access MCU/soft-switch resources.

Videoconferencing applications typically use some fixed (e.g., 1719, 1720) and some dynamic ports (port range: $> 2^{10}$ and $< 2^{15}$) to allow voice and video traffic in-and-out of the internal-network’s ports during active sessions.

Given the increasing rate of cyber-attacks on networks today, firewall policies that require opening ports for videoconferencing applications increase the risk of malicious access of internal-network data by intruders through the same open ports. To cope with this risk, network planners continually face challenges that involve balancing trade-offs between network security for data and performance of voice and video. One of the major challenges network planners have to deal with corresponds to hardware limitations of firewalls that arise when handling heavy network-traffic loads under complex firewall rule-set constraints. The per-packet inspections by firewalls of address, port and message type slow down video and voice traffic and introduce undesired lags and impairments that impact end-user quality of experience (QoE) (Calyam et al., 2004). Also, other application packet-processing loads on firewalls aggravate the lag and impact QoE. Further, if firewalls are configured properly initially by considering load demands, they do not continually function as desired due to management dynamics of networks. The management dynamics result in ever-changing security policies in networks owing to:

a. patches to block newer cyber-attack patterns
b. new end-user needs
c. network upgrades.

To overcome such challenges, middleboxes such as virtual private network (VPN) gateways and application layer gateways (ALG) such as gatekeeper proxies (GP) [GNU Gatekeeper (GNU GK), 2009; Polycom VSIU, 2009] are rapidly becoming an integral part of securing videoconferencing systems. Also, several middlebox control protocols such as ITU-T Recommendation H.460 (2005), Middlebox Communications (MIDCOM) (Srisuresh et al., 2002) and simple traversal of UDP over network address translation (STUN) (Rosenberg et al., 2003) have been proposed. The GPs are more common in the internet than VPN gateways because VPN gateways are designed mainly for internal videoconferences within a single enterprise or trusted enterprises. The middlebox solutions provide effective firewall traversal for multimedia traffic and also enable EPs to interoperate in distributed environments with heterogeneous firewall traversal solutions. However, the challenge in deploying the middlebox solutions is that they can also impact QoE when handling peak multimedia loads due to their hardware limitations. Particularly, network planners need to be aware of load handling capabilities of these middleboxes and determine appropriate placement of these middleboxes to be able to cater to the multiperiod videoconferencing demands of end-users. Inadequate planning and policy misconfigurations in deploying firewalls and middleboxes could result in vulnerable networks, slow data transfers as well as multimedia performance problems.

In this paper, we study various strategies with GPs to solve the firewall traversal challenges in securing distributed videoconferencing systems. Specifically, there are four primary contributions of this paper: the first contribution is that we review state-of-the-art standards and solutions for firewall traversal and discuss their adoption pros and cons. Our review covers non-proxy solutions as well as the more widely-used GP solutions. The second contribution is that we empirically evaluate the load-handling of GP solutions under low, medium and high cross-traffic loads using subjective and objective mean opinion score (MOS) rankings (ITU-T Recommendation P.911, 1998) and objective mouth-to-ear (M2E) delay (Jiang et al., 2003) measurements. The GP solutions can be classified under two categories:

1. commodity-hardware-based
2. appliance-hardware-based.

The cross-traffic loads feature voice, video and data streams to identify the breakdown points of the GP solutions. The third contribution is that we describe the signalling-and-multimedia flow architectures and identify caveats that arise due to heterogeneous adoption of these strategies. The caveats correspond to topology configuration issues such as GP’s deployment locations that influence neighbour gatekeeper registration constraints, bandwidth allocation and user EP dialling speeds. The fourth contribution is that we leverage our empirical results, caveats identification and vast operations experience to list best-practices for deploying GPs in small-to-large scale secure videoconferencing systems. The best-practices address issues of suitability, configurability, setup complexity and level of maintenance.

The remainder of this paper is organised as follows: Section 2 presents related work on securing videoconferencing systems. Section 3 describes the strategies used with GPs to overcome the challenges in firewall traversal in videoconferencing systems. Section 4
presents our experiments to compare load-handling performance of GP solutions under varying cross-traffic loads. Section 5 describes the signalling-and-multimedia flow architectures due to heterogeneous adoption of the firewall traversal strategies. Section 6 lists the best-practices for deploying GPs in secure videoconferencing systems, and Section 7 concludes the paper.

2 Related work

Earlier work can be categorised as having two broad themes of securing videoconferencing systems. The first theme corresponds to application-level security that involves password authentication and encryption of device-to-device communications in videoconferencing systems. The second theme corresponds to network-level security aimed at firewall traversal mechanisms for device-to-device communications in videoconferencing systems.

The threat addressed in the first theme is that if a hacker is able to hack into a personal videoconference by compromising an EP’s access password or an MCU conference session’s password, the hacker can maliciously manipulate remote management controls, and also perform unauthorised video surveillance by observing/recording the information being discussed. Consequently, the hacker can undermine an EP owner’s reputation by connecting into uninvited conferences. In addition, if the information gained by unauthorised video surveillance is confidential and high-profile, the hacker can sell the sensitive information to competitors. Occurrences of such hacking on the internet have been reported (Delio, 2002) as due to weak passwords as well as transmission of ‘clear text’ passwords in older EPs that were snooped using network sniffers. Owing to such vulnerabilities, recent government regulations such as the Health Insurance Portability and Accountability Act (HIPAA) and the Sarbanes-Oxley Act of 2002 require that medical providers and financial institutions secure personal client data by encrypting electronic transmissions that also encompass videoconferencing traffic.

Security standards relating to videoconferencing systems such as ITU-T Recommendation H.235 (2003) recommend using both password and certificate-based method for authentication and encryption; the SIP (Handley et al., 1999) protocol standard recommends using a digest authentication scheme based on a cryptographic hash function, both these are succinctly illustrated in Gemmill et al. (2004). Works such as Arkko et al. (2004) have shown that Multimedia Internet Keying (MIKEY), which involves cryptographic parameter exchange in the media negotiation, is well-suited for securing multimedia exchanges. Other works such as Featherstone and Zhang (2003) have suggested application-level security mechanisms such as using a single symmetrical encryption key exchange amongst the video conference members generate extra overheads. The overheads are not desirable especially when members frequently join/leave in mobile videoconferencing environments because links in mobile networks are typically bandwidth constrained and more error-prone compared to wired networks. In Honeyman et al. (1998), an application-level security middleware involving a smartcard-based key distribution algorithm and a fast cipher was developed for the popular VIC videoconferencing application (McCanne and Jacobson, 1995). This security middleware had better strength and was easier to implement in software than the default DES encryption option of video streams provided with the VIC application.

The threat addressed in the second theme relates to the focus of the work presented in this paper, where ports opened to allow videoconferencing traffic could be compromised by hackers to gain malicious access to internal network data. Works such as Chuan et al. (2003) and Du et al. (2008) have explored enforcing network-level security by adopting VPN gateways. Here, secure VPN tunnels are established between videoconference participants using mechanisms such as Internet Protocol Security (IPsec) (Davis, 2001) and Secure Sockets Layer (SSL) (Steinberg, 2005). The IPsec mechanism requires participants to use an IPsec client configured to connect to the network where the MCU is located, which hosts the videoconference. The SSL mechanism is easier to deploy since most web-browsers have SSL capabilities built-in, and thus no special client software install as required in IPsec are necessary. However, SSL tunnels are suitable only for web-based videoconferencing systems. The authors in Du et al. (2008) use VPN tunnels by developing custom peer-to-peer SIP (P2PSIP) protocol overlay extensions. Other earlier works such as Roberts (2001), Uwex (2009), Stoeckigt (2005, 2006) and Schlatter (2006) have considered an alternate approach of using ALGs such as such as GPs (GNUGK, 2009; Polycom Video, 2009). Our work presented in this paper is unique because we comprehensively evaluate the pros and cons and also provide a quantitative comparison of the various firewall traversal solutions in terms of subjective and objective measurements. In comparison, earlier works were based on mostly qualitative experience of users when comparing the firewall traversal solutions.

3 Firewall traversal strategies

In this section, we describe the pros and cons of non-proxy strategies as well as the more widely-used GP-based strategies, and review related standards and deployment solutions for firewall traversal in videoconferencing systems.

3.1 Non-proxy strategies

3.1.1 Open without firewall

The simplest strategy is to have an ‘open’ network without having any intermediate firewalls between EPs of the end-users and application service devices such as MCUs. Although this is ideal for videoconferencing devices, it tremendously increases the risk of data security for the internal network. Also, there is an inherent assumption in this strategy that end-users are adept enough to use
PC-based firewalls and patch their operating systems regularly to keep up with the latest cyber-attack trends.

To make such a strategy more practical to implement, data and multimedia traffic can be separated using separate VLANs. Here, the multimedia traffic can be made to bypass any firewall inspection, whereas the data traffic will undergo firewall policy checks. Although this solution reduces the risk of data security, it is not scalable for a large number of videoconferencing devices. This is because, it requires preconfiguration of the EPs and their physical locations, which introduces management overhead when newer devices are added or when there is a need for mobility of the devices. Besides, this strategy assumes that there are no firewalls along the downstream network towards a remote EP, which is not a valid assumption in today's internet.

3.1.2 EP behind firewall

Another strategy is to use 'static' configurations on firewalls that require opening predetermined set of ports for known videoconferencing devices including EPs. The advantages are that any firewall can be used for implementing this strategy and the risk of data security is reduced. However, similar to the ‘open’ case, this introduces management overhead when newer devices are added. In addition, all the videoconferencing devices need to be configured to use only a certain set of ports, which does not scale well given the proprietary issues in vendor devices and the various types of videoconferencing related services that need to be supported.

Some of the management overhead limitations can be overcome using firewalls that use ‘dynamic’ configuration options to open ports using techniques such as stateful-packet inspection (Roberts, 2001; Uwex, 2009). For example, the Cisco PIX firewall appliance uses a H.323 fixup feature with stateful-packet inspection (Uwex, 2009). However, such firewalls that support dynamic configuration options are specialised and are relatively expensive. Further, they need to be extensively tested and qualified after every software upgrade or every major firewall policy change.

3.1.3 EP in demilitarised zone alongside firewall

To isolate the data and multimedia traffic, a commonly used strategy in smaller institutions is to place the EPs in a demilitarised zone (DMZ) alongside a firewall as shown in Figure 1. This strategy does not require purchasing any specialised firewalls. However, it requires setting up a DMZ and thus requires skilled networking expertise. Also, it requires a special DMZ-connected location for videoconferencing. Such a requirement does not permit end-users to videoconference from their desktops and requires them to physically walk-over to the special DMZ-connected location for any videoconferencing.

Figure 1  EP in DMZ alongside firewall (see online version for colours)

3.2 Proxy-based strategies

3.2.1 GP in DMZ alongside firewall

Using a GP that can be placed in a DMZ alongside a firewall as shown in Figure 2 is a strategy that is widely used (Stoeckigt, 2005). The advantages of this strategy are that:

a the data security is not compromised to support multimedia requirements

b the EPs need not be located in the DMZ.

Hence, end-users can videoconference from their desktops. However, this strategy requires purchasing a GP device and introduces the overhead of maintaining it. Also, the GP device needs to be robust against cyber-attacks. Otherwise, it may be compromised by hackers and thus provides a backdoor for the hackers to gain access to the data and system resources within the internal network.

Figure 2  GP in DMZ alongside firewall (see online version for colours)
3.2.2 Standalone GP with integrated firewall

Instead of maintaining a separate GP for multimedia traffic and a firewall for data traffic, both these functions can be combined into a standalone gatekeeper proxy with integrated firewall (GP-FW) as shown in Figure 3. The advantages of this popular strategy are that:

a) security of data is not compromised to support multimedia requirements
b) there is no need for DMZ setup; hence, end-users can videoconference from their desktop
c) GP is less vulnerable to cyber-attacks due to its integration within the firewall.

In addition, this strategy provides the flexibility of retaining any predeployed firewall(s) when the GP-FW is placed in parallel to augment existing security policies with minimal disruption. However, this strategy requires purchasing a GP-FW device that is relatively expensive and also introduces the overhead of device maintenance.

Figure 3  Standalone GP-FW (see online version for colours)

3.2.3 Standards and solutions

Amongst the middlebox control protocols such as ITU-T Recommendation H.460 (2005), MIDCOM (Srisuresh et al., 2002) and STUN (Rosenberg et al., 2003), the H.460 standard has been widely adopted for proxy-based firewall traversal. There are two variants of the H.460 standard:

1) H.460.18 (signal proxy for H.225/H.245)
2) H.460.19 (multimedia proxy for RTP).

Compliance with H.460 allows firewall traversal for private-side EPs using a remote GP or GP-FW. Private-side EPs refer to EPs behind firewalls that block incoming ports for voice and video. For better understanding, let us consider a scenario where a private-side EP at Site-A initiates a videoconference session with remote EP behind a GP or GP-FW at Site-B. Upon session initiation, the GP or GP-FW realises that the Site-A EP is behind a blocking firewall and subsequently rewrites all the signalling addresses to the (detected) public IP on the Site-A firewall. More recently, new variants of the H.460 standard, i.e., H.460.23 and H.460.24 have been published that can be used in networks that do not wish to deploy a middlebox for media flows. The new variants enable direct point-to-point media between EPs, even if both the devices are behind firewall. However, for this to work, a trusted GP has to be able to negotiate a strategy to allow media flows through the respective network firewalls. A key advantage of the H.460.23 and H.460.24 recommendations is that the media flow negotiation strategy can be setup in advance, and thus the call establishment time can be reduced by avoiding the negotiation step for every call. Note that the H.460 standard requires both the EPs to be capable of H.460 signalling. Also, it requires the Site-B EP to send keep-alive messages periodically (default: 30 seconds) to keep the firewall ports at Site-A open for the session. Further, the outgoing ports shown in Table 1 must be open at the firewall at Site-A for H.460 implementations.

Table 1  Ports to be opened for H.460 support

<table>
<thead>
<tr>
<th>H.323 protocol</th>
<th>Transport protocol</th>
<th>Port numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAS</td>
<td>UDP</td>
<td>1,719</td>
</tr>
<tr>
<td>Q.931 (H.225)</td>
<td>TCP</td>
<td>1,720</td>
</tr>
<tr>
<td>H.245</td>
<td>TCP</td>
<td>14,085:15,084</td>
</tr>
<tr>
<td>RTP</td>
<td>UDP</td>
<td>16,386:34,386</td>
</tr>
</tbody>
</table>

Given the benefits and added security control in adopting proxy-based strategies, there are several commercial and open-source solutions that implement the H.460 standards. They can be categorised as either GPs/GP-FWs that are commodity-hardware-based or GPs/GP-FWs that are appliance-hardware-based. The choice of a firewall traversal solution in a deployment depends on the scale of the videoconferencing system. The scale of a videoconferencing system is determined by the maximum amount of multimedia traffic load due to concurrent videoconference sessions. Small-scale systems have < 5 Mbps loads, medium-scale systems are characterised by 5–25 Mbps loads and large-scale systems are characterised by > 25 Mbps loads. The Polycom V2IU (2009) solution is a commercially available appliance whose hardware features an application-specific integrated circuits (ASIC). A low-end appliance variant of this solution is the V2IU 4300 that can be deployed in small scale videoconferencing systems. Another appliance variant is the V2IU 5300 that can be deployed in medium-to-large scale videoconferencing systems. The GNUGK (2009) solution is the popular open-source software that can be installed on commodity-server hardware with secured and hardened Linux OS (Puschitz, 2009). The GNUGK solution can be typically deployed in small-to-medium scale videoconferencing systems. There have been several GNUGK success stories in Internet2 and ViDeNet H.323...
Evaluation of gatekeeper proxies for firewall traversal in secure videoconferencing systems

4 Gatekeeper proxy load-handling experiments

In this section, we describe our experiments to empirically evaluate the load-handling of appliance-hardware-based and commodity-hardware-based gatekeeper proxy solutions in small-to-large scale secure videoconferencing systems. We first describe our testbed setup for the experiments. Next, we explain the subjective and objective MOS rankings (ITU-T T Recommendation P.911, 1998) and objective M2E delay (Jiang et al., 2003) metrics used in our experiments. Lastly, we present our experiment results for cross-traffic loads featuring voice, video and data streams to identify the breakdown points of the gatekeeper proxy solutions.

4.1 Testbed setup

The appliance-hardware-based solutions featured in our testbed were a V2IU 4300 and a V2IU 5300, and the commodity-hardware-based solution featured in our testbed was a GNUGK. Thus, our testing covered solutions suitable for small, small-to-medium, as well as medium-to-large scale videoconferencing deployments. To compare performance of these firewall traversal solutions at corresponding cross-traffic loads, three sets of test cases were designed. The first set corresponded to the V2IU 4300 configured as a GP-FW. The goal of this testing was to evaluate the maximum data throughput performance of an appliance-based GP-FW in a small scale videoconferencing system. The second set corresponded to the GNUGK configured as a GP on commodity-server hardware. Specifically, GNUGK was installed on a typical modern 1U server with the following hardware specs: 2.4 GHz, 1 GB RAM, 80 GB Hard disk and 100 Mbps NIC. The goal of this testing was to evaluate the maximum call load performance of a commodity-hardware GP-FW proxy in a small-to-medium scale videoconferencing system. The third set corresponded to the V2IU 5300 configured as a GP-FW. The goal of this testing was to evaluate the maximum data throughput performance of an appliance-based GP-FW in a medium-to-large scale videoconferencing system.

4.2 Evaluation metrics

To evaluate multimedia degradation of the test videoconference session due to a DUT under different cross-traffic loads, we used three metrics that complement each other and provide a diverse set of performance perspectives. The first metric corresponds to the ‘subjective MOS’ metric. It is obtained by conducting human subject experiments, where each human subject quantifies his/her QoE using a ‘MOS’ ranking scheme (ITU-T Recommendation P.911, 1998). As shown in Figure 5, MOS is obtained from a human subject for a videoconference session using a five-point scale that has a mapping to the different grades of the end-user satisfaction. Specifically, the [4, 5] MOS range corresponds to ‘good’ grade where an end-user perceives none or minimal impairments and the videoconference session is always usable. The MOS [3, 4] range corresponds to ‘acceptable’ grade where an end-user
perceives intermittent impairments yet the videoconference session is mostly usable. Lastly, the [1, 3] MOS range corresponds to ‘poor’ grade where an end-user perceives severe and frequent impairments that make the videoconference session unusable. In accordance with the ITU-T recommendation to use a minimum of four human subjects for statistical soundness (Mohamed and Rubino, 2002), we had four human subjects in our subjective testing. We also used an ‘objective MOS’ metric that has the same five-point scale as the subjective MOS metric. The objective MOS was obtained using the popular NTIA’s VQM tool (Pinson and Wolf, 2004). The NTIA VQM tool is based on the ITU-T Recommendation J.144 (2001) for video quality impairment estimation. It performs instantaneous peak-signal-to-noise (PSNR) calculations using frame-by-frame comparisons of original and reconstructed video frames. Finally, it averages the instantaneous PSNR values and outputs PSNR-mapped-MOS measurements shown in Table 4. Our aim in obtaining the objective MOS results was to augment the subjective MOS results, and help us in the verification of the validity and repeatability of the subjective MOS results.

To further bolster the repeatability of MOS results in our experiments and to gain additional performance perspectives, we used a third metric in the load testing viz., ‘M2E delay’ (Jiang et al., 2003) whose range is on the order of milliseconds. We measured the M2E delay for the test videoconference session using a pulse generator connected to the test HD EP on the internal side of the testbed as shown in Figure 4. Oscilloscopes on both the open and internal sides of the testbed were calibrated such that the pulse delays due to device processing delays between the test HD EPs were measurable. The M2E metric thus indicates the lag introduced by the DUT due to device processing delays under the different load conditions. Figure 6 shows an example screenshot of an oscilloscope showing 240 ms lag in the side-B (open-side) trace due to device processing delays for each pulse in a series of pulse ticks from a pulse generator whose output is shown in the side-A (internal-side) trace.

To further bolster the repeatability of MOS results in our experiments and to gain additional performance perspectives, we used a third metric in the load testing viz., ‘M2E delay’ (Jiang et al., 2003) whose range is on the order of milliseconds. We measured the M2E delay for the test videoconference session using a pulse generator connected to the test HD EP on the internal side of the testbed as shown in Figure 4. Oscilloscopes on both the open and internal sides of the testbed were calibrated such that the pulse delays due to device processing delays between the test HD EPs were measurable. The M2E metric thus indicates the lag introduced by the DUT due to device processing delays under the different load conditions. Figure 6 shows an example screenshot of an oscilloscope showing 240 ms lag in the side-B (open-side) trace due to device processing delays for each pulse in a series of pulse ticks from a pulse generator whose output is shown in the side-A (internal-side) trace.

**Figure 5** MOS rankings mapped to MOS grades

<table>
<thead>
<tr>
<th>USER SATISFACTION</th>
<th>MOS</th>
<th>GRADE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Satisfied</td>
<td>5.0</td>
<td>Good</td>
</tr>
<tr>
<td>Satisfied</td>
<td>4.3</td>
<td></td>
</tr>
<tr>
<td>Some Users Dissatisfied</td>
<td>4.0</td>
<td>Acceptable</td>
</tr>
<tr>
<td>Many Users Dissatisfied</td>
<td>3.6</td>
<td>Acceptable</td>
</tr>
<tr>
<td>Nearly All Users Dissatisfied</td>
<td>2.6</td>
<td>Poor</td>
</tr>
<tr>
<td>Not Recommended</td>
<td>1.0</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2** PSNR mapping to MOS rankings

<table>
<thead>
<tr>
<th>PSNR (dB)</th>
<th>MOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; 37</td>
<td>5</td>
</tr>
<tr>
<td>31–37</td>
<td>4</td>
</tr>
<tr>
<td>25–31</td>
<td>3</td>
</tr>
<tr>
<td>20–25</td>
<td>2</td>
</tr>
<tr>
<td>&lt; 20</td>
<td>1</td>
</tr>
</tbody>
</table>

**4.3 Experimental results**

Figure 7 shows the MOS results of the test videoconference session for the V2IU 4300 under low, medium and high data cross-traffic loads that correspond to Iperf UDP streams of 15 Mbps, 40 Mbps and 70 Mbps, respectively. We can observe that both the subjective and objective MOS measurements in the case of the switch (without intermediate V2IU 4300) are not affected even at high cross-traffic loads. However, the V2IU 4300 shows notable degradation in MOS in the ranges of ‘acceptable’ and ‘poor’ grades for cross-traffic loads greater than 30 Mbps due to its processing power limitations.
Figure 8 shows the M2E delay measurements of the test videoconference session for the V^2IU 4300 under low, medium and high cross-traffic loads ranging from 0–70 Mbps. We can observe that the M2E delay measurements in the case of the switch (without intermediate V^2IU 4300) are not affected even at high cross-traffic loads. We refer the M2E delay measurement in the switch case as ‘control’. The control M2E delay value corresponds to the sum of encode and decode times of the EP codecs and the switch processing delay. In comparison to the switch case, the V^2IU 4300 shows notable degradation in M2E delay performance for cross-traffic loads greater than 30 Mbps due to its processing power limitations (i.e., M2E delay results are consistent with MOS results). The V^2IU 4300 causes noticeable fluctuations with low jitter-levels till about 50 Mbps. For traffic loads greater than 50 Mbps, the V^2IU 4300 causes the output pulse signal jitter to be highly variable with peaks greater than 300 ms. Note that earlier studies such as ITU-T Recommendation G.114 (1996) and Calyam et al. (2004) have shown that such peak M2E delay values seriously hamper interactive communications between end-users.

Figure 9 shows the MOS results of the test videoconference session for the GNUGK under varying cross-traffic loads. Note that this set of test cases for GNUGK involved video and voice cross-traffic loads up to 15 Mbps only. This is because GNUGK being a video proxy device can be loaded only using videoconferencing cross-traffic and thus data cross-traffic loads of Iperf UDP streams are not relevant here. Based on the availability of SD and HD EPs in our testbed, we were able to generate only up to 15 Mbps of videoconferencing cross-traffic loads. Nevertheless, we can observe that both the subjective and objective MOS measurements in the case of the switch and with intermediate GNUGK show negligible degradation for 15 Mbps cross-traffic loads. Hence, we can conclude that the GNUGK running on typical modern 1U server hardware may handle up to 15 Mbps of cross-traffic loads well, and higher cross-traffic loads will expose the processing power limitations.

Figure 11 shows the MOS results of the test videoconference session for the V^2IU 5300 under low, medium and high data cross-traffic loads that correspond to Iperf UDP streams of 15 Mbps, 40 Mbps and 70 Mbps, respectively. We can observe that both the subjective and objective MOS measurements in the switch case and in the intermediate V^2IU 5300 case show negligible degradation even at high cross-traffic loads. Hence, we can conclude that the V^2IU 5300 maintains ‘good’ grade MOS even for high cross-traffic loads due to its relatively superior processing power capabilities.
Figure 11 MOS for V2IU 5300 under varying loads (see online version for colours)

Figure 12 shows the M2E delay of the test videoconference session for the V2IU 5300 under low, medium and high cross-traffic loads ranging from 0 - 70 Mbps. We can observe that both the M2E delay measurements in the case of the switch and with intermediate V2IU 5300 are not affected even at high cross-traffic loads and thus the M2E delay results confirm the conclusions drawn from the MOS results for the high cross-traffic loads.

Figure 12 M2E delay for V2IU 5300 under varying loads

5 Signalling-and-multimedia flow architectures

In this section, we describe the signalling-and-multimedia flow architectures and identify caveats that arise due to heterogeneous adoption of the various firewall traversal strategies.

Depending upon the topology configuration and the extent of heterogeneous adoption of the various firewall traversal strategies, different signalling-and-multimedia flow architectures occur between EPs. To better understand the flow architectures, let us consider Figure 13 that shows an example scenario with different EP characteristics shown in Table 3. To be able to communicate between any two EPs, two sets of signalling messages need to be exchanged as marked in Figure 13. The first set corresponds to the EP registration messages with the GPs or GP-FWs. The second set corresponds to neighbouring GKs, i.e., GPs or GP-FWs. After exchanging such signalling messages, EP-5 can initiate a session to EP-3 because GP-FW-1 knows GP-FW-2 and GP-FW-1 knows GP-DMZ. Table 4 shows the details of the signalling and multimedia flow paths for communicating between any two EPs in the example scenario.

Table 3 EP configurations in example scenario

<table>
<thead>
<tr>
<th>EP</th>
<th>Configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>EP-1</td>
<td>EP behind firewall</td>
</tr>
<tr>
<td>EP-2</td>
<td>EP in DMZ</td>
</tr>
<tr>
<td>EP-3</td>
<td>EP inside private network with GP in DMZ and 3rd party firewall</td>
</tr>
<tr>
<td>EP-4</td>
<td>EP inside private network with GP-FW</td>
</tr>
<tr>
<td>EP-5</td>
<td>EP inside private network with GP-FW</td>
</tr>
</tbody>
</table>

Figure 13 Example scenario showing signalling and multimedia flow patterns (see online version for colours)

Given the multimedia flow paths in different topology configurations, there could be bottleneck issues with intermediate GPs or GP-FWs handling heavy loads of traffic. The traffic may not only belong to the internal-network end-users but also may correspond to traffic from external end-users whose GPs or EPs have neighboured or registered with the local GP or GP-FW. We can also note that the worst case scenario can have multimedia between EPs passing through two GPs or GP-FWs at the most. Nevertheless, careful planning must be undertaken for the GP and GP-FW placement and registration, particularly in large-scale and hierarchical videoconferencing systems. In such networks, improper GP and GP-FW placement and registration could not only cause overloads on GP and GP-FW, but also cause available bandwidth bottlenecks between EPs and consequent constrained EP dialling speeds that degrade multimedia performance and data throughputs.
Table 4  Signalling and multimedia flow paths between EPs of example scenario

<table>
<thead>
<tr>
<th>EP-pair</th>
<th>Signalling flow path</th>
<th>Multimedia flow path</th>
</tr>
</thead>
<tbody>
<tr>
<td>EP-1 ↔ EP-2</td>
<td>GP-FW-1</td>
<td>None</td>
</tr>
</tbody>
</table>

6  Best practices for gatekeeper proxy deployment

In this section, we leverage our empirical results, caveats identification and vast operations experience to list best-practices for deploying gatekeeper proxies in small-to-large scale secure videoconferencing systems.

Table 5  Comparison of firewall traversal solutions considering DMZ and proxy/firewall requirements

<table>
<thead>
<tr>
<th>Traversal solution</th>
<th>DMZ requirement</th>
<th>Proxy/firewall requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cisco PIX with H.323 fixup</td>
<td>No</td>
<td>No proxy required; device is an H.323 protocol aware firewall</td>
</tr>
<tr>
<td>Polycom V2IU 4350</td>
<td>Yes – is used only as proxy; No – if used as a firewall or in parallel with third party firewall</td>
<td>Device acts as a proxy and has integrated firewall; third party firewall required that needs high-maintenance and setup complexity</td>
</tr>
<tr>
<td>Polycom V2IU 5300</td>
<td>Yes – if used only as proxy; No – if used as a firewall or in parallel with third party firewall</td>
<td>Device acts as a proxy and has integrated firewall; third party firewall required that needs high-maintenance and setup complexity</td>
</tr>
<tr>
<td>GNUGK</td>
<td>Yes – only used as proxy</td>
<td>Device acts as a proxy and has integrated firewall; third party firewall required that needs high-maintenance and setup complexity</td>
</tr>
</tbody>
</table>

Table 6  Comparison of firewall traversal solutions considering price-performance trade-offs

<table>
<thead>
<tr>
<th>Traversal solution</th>
<th>Suitability</th>
<th>Setup complexity</th>
<th>Level of maintenance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cisco PIX with H.323 fixup</td>
<td>Enterprise and ISP (can sustain multimedia plus data loads up to 70 Mbps and possibly beyond)</td>
<td>High (requires skilled engineering expertise)</td>
<td>High (software upgrades, testing after major rule updates)</td>
</tr>
<tr>
<td>Polycom V2IU 4350</td>
<td>Enterprise (cannot sustain multimedia plus data loads beyond 30 Mbps)</td>
<td>Low (requires videoconferencing administrator expertise)</td>
<td>Low (software upgrades, extensive testing to verify upgrade done by vendor)</td>
</tr>
<tr>
<td>Polycom V2IU 5300</td>
<td>Enterprise and ISP (can sustain multimedia plus data loads up to 70 Mbps and possibly beyond)</td>
<td>Low (requires videoconferencing administrator expertise)</td>
<td>Low (software upgrades, extensive testing to verify upgrade done by vendor)</td>
</tr>
<tr>
<td>GNUGK</td>
<td>Enterprise (can sustain multimedia loads up to 15 Mbps, and may not sustain multimedia loads more than 20 Mbps; highly dependent on the device hardware)</td>
<td>Medium (requires skilled system administrator for OS hardening, and videoconferencing administrator)</td>
<td>High (software upgrades of GNUGK and Linux OS, extensive testing after upgrades)</td>
</tr>
</tbody>
</table>
Have fail-over options: In any network or videoconferencing operations environment, spares and fail-over strategies are critical to avoid service outages that annoy end-users. Hence, it is pertinent to have one or more extra firewall traversal solutions deployed or ready to be deployed for redundancy purposes.

Use a dedicated GNUGK device: If deploying the GNUGK proxy, ensure a dedicated server is available for the GNUGK software and turn-off other system services (e.g., web server) that may compete for system and network resources. Also, it is important to keep the Linux OS on the GNUGK proxy server secured and hardened by regularly applying security patches.

Test and document: Regularly test to ensure compatibility of EPs, MCUs and GP or GP-FW with the H.460 protocols, especially in a multivendor videoconferencing system. Also, periodically test the traffic load patterns and perform suitable load balancing to tune performance. Testing is also recommended when major firewall policy changes are made, including software updates and firewall rule-set modifications.

7 Conclusions

For the past several years, there has been a tension between network security engineers configuring firewalls to ensure data security and videoconferencing engineers configuring videoconferencing applications to ensure high quality voice and video. Given the wide variety of network designs, reproducing usability problems relating to firewalls, affecting voice and video user QoE performance, have always been challenging and often frustrating. With the recent emergence of new standards such as ITU-T H.460 and related firewall traversal gatekeeper proxy solutions, there are now several options to balance the trade-offs between ensuring network security for data, and superior QoE performance of voice and video. In this paper, we studied non-proxy and gatekeeper proxy-based strategies to solve the firewall traversal challenges in securing distributed videoconferencing systems. Using systematic experiments, we evaluated the load-handling of commodity-hardware-based and appliance-hardware-based gatekeeper proxy solutions under low, medium and high cross-traffic loads using subjective and objective measurements. Further, we described the signalling-and-multimedia flow architectures and identified caveats that arise due to heterogeneous adoption of the various firewall traversal strategies. The caveats correspond to topology configuration issues such as gatekeeper proxy’s deployment locations that influence neighbour gatekeeper registration constraints, bandwidth allocation and user EP dialling speeds. Lastly, by leveraging our empirical results, caveats identification and vast operations experience, we listed best-practices for deploying gatekeeper proxies in small-to-large scale secure videoconferencing systems. The best-practices addressed issues of suitability, configurability, setup complexity and level of maintenance.

Acknowledgements

This work has been supported by the Ohio Board of Regents, OARnet and Polycom Corporation.

References


Notes

1 Most firewalls also provide network address translation (NAT) functionality that allows internal-network hosts to access the internet using a public IP address. Hence, firewall traversal is also referred to as firewall/NAT traversal in the literature.

2 Stateful-packet-inspection is a firewall feature that keeps track of out-bound packets and associates in-bound packets with hosts of out-bound packets. Thus, this feature allows safe handling of traffic without complex pre-configuration of firewall rules.