

A First Look at Media Conferencing Traffic in the Global Enterprise

Vijay Vasudevan^{1,2}, Sudipta Sengupta², and Jin Li²

¹ Carnegie Mellon University, Pittsburgh PA 15213, USA

² Microsoft Research, Redmond WA 98052, USA

Abstract. Many enterprise networks have grown far beyond a single large site to span tens to hundreds of branch offices across the globe, each connected over VPNs or leased lines. With the emergence of the globally-connected enterprise and the trend towards enterprise all-IP convergence, including transitioning the enterprise PBX to VoIP servers, IP-based audio/video conferencing for telepresence has challenged the notion that bandwidth is abundant in the enterprise. We take a first look at media conferencing traffic in the global enterprise and, by instrumenting measurement of call quality and network statistics, we quantify the impact on call quality for a range of factors in the enterprise, such as wired vs. wireless access, inter- vs. intra- branch office communication, QoS mechanisms like VLAN tagging and DiffServ DSCP marking, and VPN vs. public Internet access.

1 Introduction

Enterprise traffic analysis has received only scarce attention in the networking community, due in part to the difficulty of recording enterprise traffic. Early studies on enterprise traffic have found that network utilization is typically 1-3 orders less than network capacity, suggesting that enterprise networks have abundant bandwidth provisioning, unlike the WAN or global Internet [1]. This belief can no longer be assumed, primarily because of two trends: 1) the emergence of the globally-connected enterprise that spans countries and continents, with branch offices connected by leased lines, Virtual Private Networks (VPNs), and public Internet connections, and 2) the trend towards enterprise all-IP convergence, including transitioning the enterprise PSTN-based telephone exchange deployments to VoIP servers and IP-based audio/video conferencing for telepresence.

In particular, the move to VoIP and video conferencing is fundamentally changing enterprise traffic characteristics. Worldwide VoIP service is projected to grow to \$24.1B in 2007, up 52% from 2006. It is expected that worldwide VoIP service revenue will more than double over the next 4 years, reaching \$61.3B in 2011 [2]. To the best of our knowledge, there is no large-scale quantified measurement study of the quality of service for VoIP and video conferencing in the enterprise, despite its rapid expected growth.

In this paper, we attempt to characterize the network quality of audio/video conferencing and VoIP communication within a globally-connected and diverse

enterprise network. Using call logs from over 100,000 endpoints within this network over several months, we seek to understand several characteristics of enterprise network traffic and provisioning, including coarse statistics for packet loss, bandwidth utilization, and the causes of and potential solutions for poor call quality.

Our main results can be summarized as follows:

- Endpoints on wireless experience poor call quality as a result of a combination of non-negligible packet loss rates and high packet burstiness.
- Users connected to the enterprise over VPN or from home connections experience significantly higher packet loss rates.
- Non-wireless calls with DiffServ QoS priority treatment very rarely experience packet loss rates $> 2\%$, a loss rate which many audio codecs can mask.

We believe these results have implications for both enterprise network and protocol/application design, and suggest that large enterprise networks may not always be sufficiently provisioned given the growth of media conferencing traffic in the enterprise.

2 The Globally-Connected Modern Enterprise

Unified communication, which uses an IP-based network to carry many forms of communication traffic (e.g. VoIP, video conferencing, bulk data, and instant messaging) is gaining momentum in the industry. Most large enterprises today are turning their attention to unified communication to improve business processes, reduce communication delays, and improve employee productivity. While small to medium-sized enterprise networks consist of a small set of LANs within a single site [1], large, modern enterprise networks have grown far beyond this size, comprising several IP devices per employee, hundreds to thousands of employees per site, and tens to hundreds of sites around the world.

To accommodate this increasingly diverse geographic distribution, the scope of the modern enterprise has expanded to form a very heterogeneously-connected network. Using a large IT company with hundreds of branch offices across the globe as an example, our study focuses on the deployment and growing use of IP-based telephony and conferencing within this class of enterprise networks.

This network comprises the main campus networks located in the U.S., China, and Singapore via well-provisioned private lines (e.g. OC-48) and includes a large set of branch office sites, connected via VPN over the public Internet or over leased lines, with bandwidth capacity ranging from 1.5Mbps to 45Mbps depending on the size of the branch office. As a result, the modern large enterprise differs significantly from typical single-site enterprise deployments. Even within a single site, devices will use wireless or wired connectivity. Although the network is managed by one administrative body, the distribution of branch offices makes it difficult to provision all offices systematically to meet (variable) traffic demands. As a result, the assumption that enterprise networks are likely to have sufficient bandwidth throughout no longer holds.

As a step towards understanding such large-scale enterprise networks, in this paper we try to characterize one aspect of traffic within the network: IP telephony and audio/video conferencing, a growing portion of traffic that requires strict guarantees on performance for effective communication. While still preliminary in scope, we hope this investigation provides another initial step towards understanding the use and deployment of enterprise networks and motivates further study in the area.

3 Measurement Methodology and Datasets

3.1 Measurement

Our datasets consist of several IP phone and audio/video conference call log databases. A call can span multiple users at multiple branch offices. Each branch office has one or more *Media Control Unit (MCU)* servers whose function is to bridge the call – an audio/video stream originating at a user endpoint (client) is received at an MCU and replicated to all other participants (the choice of the MCU location for a given call depends on the location of the earliest joining user in the call). Thus, a call consists of multiple *sessions*, each session identified by a {user endpoint, MCU} pair. For a two-party call, the MCU is not involved, hence the session identified by a {user endpoint, user endpoint} pair. If any of the conference participants is on the PSTN (Public Switched Telephone Network), then a PSTN-gateway server is also involved.

For every IP phone call or audio/video conference call made within the enterprise, a *session-level* report is logged at the end of the call by participating endpoints to a central reporting server. Each individual session log contains several pieces of information for streams in both directions, including start and end timestamps, average/max packet loss, latency, average/max jitter, bandwidth capacity estimate, network quality (also known as *Network Mean Opinion*

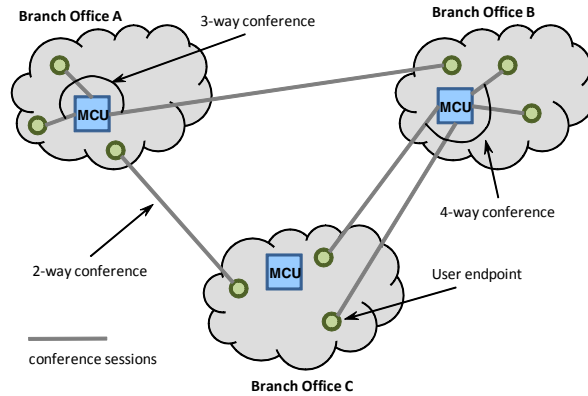


Fig. 1. Illustrating sessions in multi-party audio/video conference calls across branch offices – the session is the level at which measurements are logged

Score, or *NMOS*), device type, audio/video codec used, bitrate, etc. While the dataset does not include packet-level traces, these call log summary statistics are useful for identifying the cause of poor call quality in the enterprise, and allow for a first-look at new kinds of modern enterprise traffic characteristics. Although packet-level traces would allow for detailed traffic analysis, we instead focus on a broad characterization of call traffic in the enterprise by analyzing call volume, the effect of QoS, the real-world impact of wireless on call quality, and individual call statistics to identify the growth and potential implications for future enterprise network design.

3.2 Datasets

Table 1 describes two datasets we obtained for our characterization study. Dataset D_1 consists of a log set for a “dogfood” test deployment within the enterprise, while D_2 contains a recent, full log database of all calls. Because our logs comprise usage data for over 100,000 endpoints, our study focuses on a broad characterization of the pervasive IP-based conferencing traffic statistics in the enterprise network. We note that the number of distinct endpoints is larger than the number of users in the enterprise, because many users make calls from several endpoints (e.g., desktop machine, wireless laptop, IP phone). We use the D_2 dataset for all our analysis due to its increased log density and recency, but mention D_1 here only to note that the results obtained from both datasets are similar (except for call density), suggesting that the results presented in this paper cannot simply be attributed to initial deployment and provisioning hurdles.

Table 1. Overall Dataset Characteristics

| Data | Start Date | End Date | #Audio Streams | #Video Streams | Distinct IP Endpoints |
|-------|------------|----------|----------------|----------------|-----------------------|
| D_1 | 09/12/07 | 01/16/08 | 532,191 | 49,235 | 17,118 |
| D_2 | 01/21/08 | 06/22/08 | 9,744,660 | 617,018 | 205,526 |

We supplement our datasets with location information pulled from an internal database that contains the location of all IP subnets within the enterprise network. More specifically, the location information is of the form {IP Subnet/mask, Country, City, Building}. For each IP address in our dataset, we perform a longest-prefix match to map the location of the caller to a specific building, city, and country. For security and privacy reasons, however, we do not obtain router-level topology information for the enterprise network. In total, the location database contains entries for over 7,000 subnets spanning 500 unique buildings/locations around the world.

4 Data Analysis

In analyzing our datasets, we break down the results into two categories: trends and statistics, and quality diagnosis. The first category discusses overall media

traffic characteristics within the enterprise, such as aggregate call patterns, and overall traffic trends. The second category concerns evaluation of call quality and diagnosis of aspects that can impact call quality, such as QoS mechanisms like DiffServ [3], wired vs. (last-hop) wireless usage, and integration with the public Internet.

4.1 Trends and Statistics

Growth Trends: In Figure 2, we plot the number of audio and video streams per day over a six month period from D_3 . During this period, total call volume nearly tripled as a result of an aggressive IP phone deployment across the enterprise starting in April 2008. On the other hand, the number of video streams did not increase as dramatically over the six month span, showing 60% growth over six months. We note, however, that the more gradual video stream growth is not connected to any particular external deployment effort and represents a trend towards increased video conferencing.

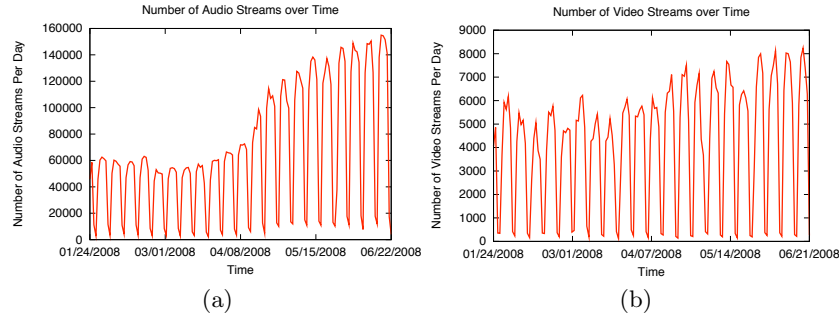


Fig. 2. (a) Audio trends follow call trends closely; growth in call volume was mostly due to increased VoIP and audio conferencing traffic. (b) Video calls increase slower in frequency in comparison to audio calls.

Call Patterns: We note several types of patterns in our dataset, ranging from time-of-day and time-of-week effects, to conference call patterns. Figure 3(a) shows the typical media traffic patterns observed at intermediate media devices (PSTN-gateways or conference stream aggregators) over a representative week. As expected, traffic exhibits a strong time-of-week effect, with peak usage during mid-week and a strong drop-off on weekends. Figure 3(b) depicts the same data for a typical Wednesday, with traffic always seen between business hours and occasional peaks due to individual large conferences. We also note the existence of many larger conference calls that occur at weekly intervals, predictably producing media traffic peaks.

4.2 Call Quality Diagnosis

Feedback indicating poor quality calls motivated us to investigate the cause of such calls and the impact of some efforts to improve call quality. Several factors

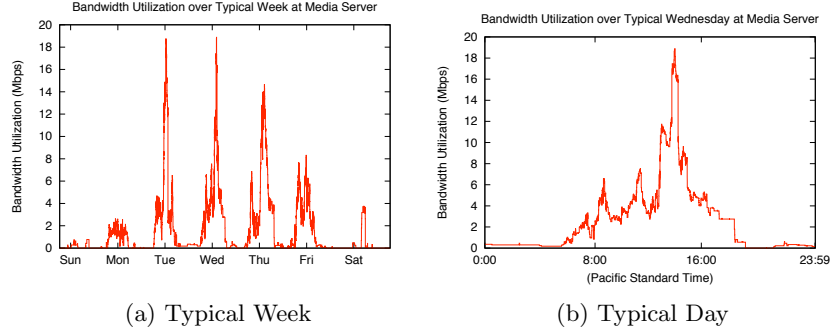


Fig. 3. Call traffic seen at a typical media aggregator

lead to poor call quality, including network-specific factors like packet loss, jitter, delay, but also application-specific factors such as audio echo, codec bitrate and frame loss rates. While Mean Opinion Scores (MOS) do help differentiate poor calls from acceptable ones, we find that max packet loss is one important contributor to poor calls from the network standpoint. In this dataset, packet loss is calculated over time intervals of 30 seconds during the call, with the maximum packet loss interval logged at the end of the call. Max packet loss is a good indicator of a poor call because even a single moment of high packet loss can distract callers and lead to a poor opinion of call quality. To this end, we focus on identifying the cause of, and potential remedies for high packet loss.

Wireless Last-Hop: While enterprise networks are often comprised of desktop and server machines on wired LANs, a significant number of hosts in the large enterprise we study actively use wireless LANs, a trend that is likely to increase given advancements in wireless deployments and ease of use. Much research has focused on the interplay between VoIP and 802.11-based wireless access, noting poor VoIP call quality on these wireless deployments despite having sufficient bandwidth capacity [4, 5]. In Figure 4(a), we plot the CDF of max packet loss, distinguishing between streams with endpoints communicating over wireless last hops. Streams involving wireless experience much worse packet loss characteristics than those on wired connections. For example, 5-10% of streams experience packet loss rates above 4%, producing poor audio quality and yielding a very poor user experience for those using video conferencing on these wireless devices. Although calls where both endpoints are wireless tend to be worse than with just one wireless endpoint, the marked difference is between wireless and non-wireless calls.

Coarse packet loss statistics alone do not tell the entire story. Most telephony and conferencing protocols are also sensitive to the duration of packet loss bursts, the length of time for which a large fraction of packets are discarded because of delayed arrival. As shown in Figure 4(b), audio streams with at least one wireless endpoint experience more and longer burst durations on average than wired connections. Thus, even if the wireless streams do not drop packets, the packets that do arrive may be useless to the real-time application.

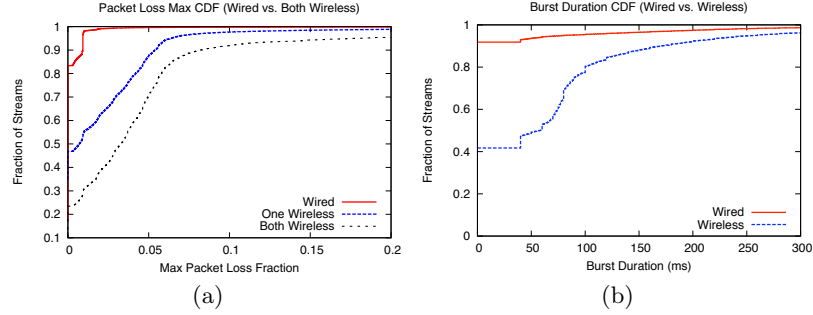


Fig. 4. CDF of packet loss fraction and burst duration broken down based on whether endpoints have a wireless last hop

Given the prevalence of (IEEE 802.11-based) wireless users (e.g., those with laptops in conference rooms), these real-world results motivate the need for deployment of newer VoIP and conferencing protocols that can accommodate such high-loss rates and minimize burst without a large degradation in call quality.

Home Users and VPN Clients: While calls within the enterprise experience higher than expected packet loss rates, we compare intra-enterprise calls to those calls that transit part of the public Internet to better understand whether the issue is specific to this enterprise network. Figure 5 plots the CDF of packet loss, separating in-enterprise endpoints and VPN/home endpoints. While packet loss within the enterprise is higher than expected, packet loss for VPN and home users is very high: 5% of streams experience packet loss rates greater than 10%. This suggests that the enterprise network is indeed better provisioned than the public Internet, though may not be as underutilized as previously noted [1]. We also observe that for external users, the call quality is perceivably better over the public Internet than over VPN. For example, an additional 10% of VPN streams experience packet loss of 5% or higher than public Internet streams. This can be attributed to VPN servers not adequately handling voice traffic, since they were designed with email/corporate-intranet traffic in mind.

Without packet-level traces or other data-traffic originating from or terminating at those endpoints, it is difficult to conjecture why these rates are much higher and more prevalent across the public Internet. For example, these higher loss rates may be caused by network congestion at the endpoint's ISP, wireless packet loss at the home user's network, or local congestion with other bulk data traffic from the endpoint itself.

QoS and Voice VLAN Usage: Most IP phone deployments are configured to exist on a separate voice VLAN [6]. Packets sent from the IP phone are marked with prioritized DiffServ DSCP [3] bits. Core routers within the enterprise and outgoing inter-branch office interfaces prioritize this voice VLAN traffic over normal best-effort data traffic. In Figure 6(a), we plot the CDF of the max packet loss fraction broken down based on whether zero, one, or both endpoints exist on

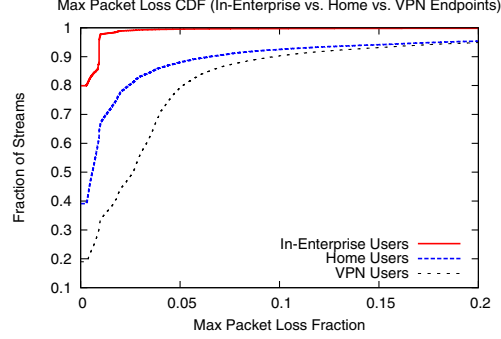


Fig. 5. Maximum packet loss CDF depending on type of user: within enterprise, external over public Internet (home), and external over VPN

the voice VLAN. In this graph, we remove calls that involve wireless endpoints to focus on the benefits of QoS on the wired LAN. Further, we remove endpoints outside the enterprise network (e.g., calls from employee homes, through VPNs) to avoid paths that traverse the public Internet, which does not respect DiffServ priority.

When both endpoints are on the voice VLAN, nearly all calls experience less than 2% packet loss, a loss rate that most audio codecs can accommodate. When exactly one endpoint is on the voice VLAN, max packet loss increases: 25% of audio streams exhibit some non-negligible packet loss. However, most calls do not experience more than 4% packet loss. When neither endpoint is on the voice VLAN, we find that 2% of streams experience packet loss rates greater than 4%. These results suggest that QoS for voice traffic can mostly eliminate the prevalence of poor audio calls due to packet loss in the wired enterprise.

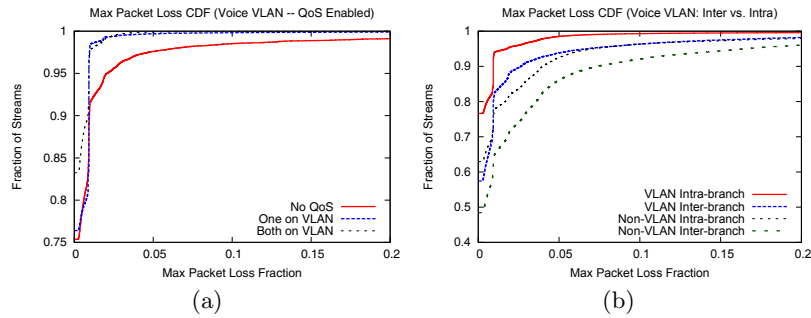


Fig. 6. (a) CDF of max packet loss fraction broken down into endpoint QoS support. Calls that travel on the voice VLAN are DiffServ-enabled and prioritized by core routers and at inter-branch routers. (b) Max packet loss broken down based on whether call involves VLAN and whether call traverses inter-branch office links.

Because QoS is more likely to help on congested paths, we further break down packet loss characteristics based on whether the call traverses an inter-office branch link or remains within the same branch office in Figure 6(b). Given the same QoS capability, packet loss rates on intra-branch paths are lower than on inter-branch paths, suggesting that inter-branch paths are likely bottlenecked at a common chokepoint. Interestingly, inter-branch office calls with QoS enabled tend to perform better than intra-branch office calls without QoS: there is enough congestion within a single site that intra-branch calls without QoS can suffer more than prioritized inter-branch calls. Thus, even the assumption that single sites are well-provisioned within an enterprise network may not be true.

5 Related Work

Measurement studies of enterprise networks have, up until recently, been scarce, despite the growth of the enterprise sector within industry. In the past decade, Shaikh et al. studied OSPF behavior in an enterprise network [7], and Pang et al. provided traffic analysis of a modern, medium-sized enterprise network, focusing on categorizing the types of traffic and network behavior seen in the enterprise based on packet traces [1]. Recent studies have also focused on enterprise network management [8], troubleshooting [9], or wireless diagnosis [10]. More recently, Guha et. al identify mobile (often off-site) hosts for degraded “health” in enterprise networks, where health is defined as the fraction of useful flows [11].

Our paper looks away from operational aspects of enterprise networks and instead attempts to shed light on an important and growing class of traffic within the modern enterprise: VoIP and conferencing. In addition, we provide a first look into a new class of modern enterprise networks that spans the globe and is heterogeneous in connectivity, making such networks more like the wide-area Internet in scope.

6 Conclusion

To the best of our knowledge, this is the first work in the literature to measure and analyze media traffic in the global enterprise. Using session-level reports logged at the end of each call containing call quality and network statistics, we quantify the impact on call quality for a range of factors in the enterprise, including wired vs. wireless access, inter- vs. intra- branch office communication, QoS mechanisms like VLAN tagging and DiffServ DSCP marking, and VPN vs. public Internet access. With the modern enterprise spanning the globe, the transitioning to VoIP from PBX, and rich media applications like voice and video conferencing seeing explosive growth in the enterprise, it can no longer be assumed that bandwidth is “abundant” in the enterprise. Our studies indicate that careful network planning and provisioning may be required in the enterprise to support good quality for media applications and that QoS mechanisms for prioritized traffic handling do indeed help to improve quality. Our continuing work in this area involves further analysis of media traffic in the enterprise using packet level traces and active measurement.

References

- [1] Pang, R., Allman, M., Bennett, M., Lee, J., Paxson, V., Tierney, B.: A First Look at Modern Enterprise Traffic. In: Proc. ACM SIGCOMM Internet Measurement Conference (October 2005)
- [2] Machowinski, M.: VoIP services and subscribers: annual market share and forecasts. In: Mobile Computing and Communications Review (2008)
- [3] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., Weiss, W.: An Architecture for Differentiated Services. Internet Engineering Task Force, RFC 2475 (December 1998)
- [4] Wang, W., Liew, S.C., Li, V.: Solutions to Performance Problems in VoIP over 802.11 Wireless LAN. IEEE Transactions on Vehicular Technology (January 2005)
- [5] Niculescu, D., Ganguly, S., Kim, K., Izmailov, R.: Performance of VoIP in a 802.11 Wireless Mesh Network. In: Proc. IEEE INFOCOM (March 2006)
- [6] IEEE 802.1Q Virtual LAN Standard,
<http://www.ieee802.org/1/pages/802.1Q.html>
- [7] Shaikh, A., Isett, C., Greenberg, A., Roughan, M., Gottlieb, J.: A Case Study of OSPF Behavior in a Large Enterprise Network. In: Proc. ACM SIGCOMM Internet Measurement Workshop (November 2002)
- [8] Casado, M., Freedman, M.J., Pettit, J., Luo, J., McKeown, N., Shenker, S.: Ethane: Taking Control of the Enterprise. In: ACM SIGCOMM 2007 (August 2007)
- [9] Bahl, P., Chandra, R., Greenberg, A., Kandula, S., Maltz, D.A., Zhang, M.: Towards Highly Reliable Enterprise Network Services Via Inference of Multi-level Dependencies. In: ACM SIGCOMM 2007 (August 2007)
- [10] Cheng, Y.C., Afanasyev, M., Verkaik, P., Benko, P., Chiang, J., Snoeren, A.C., Savage, S., Voelker, G.M.: Automating Cross-Layer Diagnosis of Enterprise Wireless Networks. In: ACM Sigcomm 2007 (August 2007)
- [11] Guha, S., Chandrashekar, J., Taft, N., Papagiannaki, K.: How Healthy are Today's Enterprise Networks? In: IMC 2008: Proceedings of the 8th ACM SIGCOMM conference on Internet measurement (2008)