Can You Hear Me Now?! It Must Be BGP

Nate Kushman
MIT - CSAIL
nkushman@mit.edu

Srikanth Kandula
MIT - CSAIL
kandula@mit.edu

Dina Katabi
MIT - CSAIL
dk@mit.edu

ABSTRACT

Industry observers expect VoIP to eventually replace most of the existing land-line telephone connections. Currently however, quality and reliability concerns largely limit VoIP usage to either personal calls on cross-domain services such as Skype and Vonage, or to single-domain services such as trunking, where a core ISP carries long-distance voice as VoIP only within its backbone, to save cost with a unified voice/data infrastructure. This paper investigates the factors that prevent cross-domain VoIP deployments from achieving the quality and reliability of existing land-line telephony (PSTN). We ran over 50,000 VoIP phone calls between 24 locations in US and Europe for a three-week period. Our results indicate that VoIP usability is hindered as much by BGP's slow convergence as network congestion. In fact, about half of the unintelligible VoIP samples in our data occur within 10 minutes of a BGP update.

Categories and Subject Descriptors

C.2.2 [**Communication Networks**]: Network Protocols; C.4 [**Performance of Systems**]: Reliability, Availability and, Serviceability

General Terms

Experimentation, Measurement, Performance, Reliability

Keywords

VoIP, BGP, Burst Losses, PESQ, MOS

1. Introduction

Recent widespread deployment of broadband Internet access has opened up the market for VoIP. Analysts estimate that 20% of voice traffic in North America will run over IP by 2010 [17]. Despite this, many business users are holding back due to quality and availability concerns [18], and cross-domain providers, such as Skype and Vonage, have not yet gained traction with their business offerings [14]. Cable companies address the performance problems by offering voice services that run as VoIP only within their own network, and transfer the calls to the PSTN as they exit the provider's network [1]. Large ISPs have used a similar technique to allow them to offer long-distance trunking services which they run as VoIP within their own network, but connect through the PSTN on either end [14]. As cable-network and ISP-based VoIP deployment increases however, it quickly reaches a point where maintaining any connection to the PSTN is cost prohibitive, and cross-domain VoIP solutions must be explored.

Past research on cross-domain VoIP, however, has not attempted to understand the root cause of the performance problems in a way that may help network operators prevent them [8, 11, 12]. It has instead focused only on documenting the performance problems, leading to a prevailing assumption that all of VoIP's performance problems are caused by congestion.

This paper, in contrast, provides convincing evidence that congestion is not the only significant problem; and in fact, almost half of the performance problems are caused by BGP. To realistically evaluate the quality of actual VoIP calls over the public wide-area Internet, we employ an open source VoIP implementation, Linphone [2], that is also used by the popular Google Talk client. We ran more than 50,000 automated phone calls over 100 paths in Europe and US, over a period of 3 weeks. We evaluate the user perceived quality of the actual voice streams using an industry standard signal processing technique, called PESQ [23], that approximates the Mean Opinion Score (MOS) of human judges. We then use BGP feeds matching the studied routes to discover whether poor call quality correlates with BGP updates. Our study reveals the following:

(1) While cross-domain VoIP is usable, its performance is not yet acceptable for a public phone network replacement. More specifically:

- VoIP is unintelligible on average for 5-10 min a day, an order of magnitude more than the public phone network [36].
- There is a 6.2% chance that a user making a long VoIP call, will hang up within the first hour due to network outages.

(2) Poor VoIP performance is highly-correlated with BGP updates. In particular:

- Surprisingly, almost 50% of the unintelligible VoIP samples occur within 10 minutes of a BGP update. This concentration is striking given that less than 1% of our overall VoIP samples are within 10 minutes of a BGP update.
- On average, when a BGP update happens, voice quality falls to an unintelligible level. Further, in more than 20% of outages correlated with BGP, the call remains unintelligible for over four minutes, preventing callers from re-establishing the call.[1] In contrast, more than 90% of

[1]This does not mean that the average BGP event lasts for 4

outages not correlated with BGP last for less than 10 seconds, readily allowing call re-establishment.

These results indicate that, contradictory to expectations, VoIP usability is affected by BGP as much as by network congestion. Correlation does not necessarily imply causality; in fact the underlying cause for poor performance is most likely a link failure or a policy change. However, BGP's inability to quickly re-converge after the event, magnifies the effect of such events on VoIP call quality. To make matters worse, while existing basic QoS and intra-domain recovery techniques (such as MPLS fast reroute) [35] can alleviate congestion and susceptibility to intra-domain link failures, improving BGP dynamics is still an open research problem.

We believe this work provides the first evidence of the correlation between poor VoIP performance and BGP updates. Our work motivates the need to improve BGP dynamics. Without fixing BGP, we leave the users with a VoIP service that is usable most of the time, but suffers from frequent outages and un-intelligibility whenever BGP's convergence comes into play. Businesses are unlikely to move to VoIP unless these availability problems are addressed [18], since a dropped call can mean a lost customer. Lastly, we believe that purely reactive techniques will not be enough. Instead, we need to augment BGP to provide a mechanism which ensures destinations remain connected during ongoing BGP convergence, as long as the underlying network is connected. [28, 46, 9].

2. Related Work

Prior work has looked at VoIP performance and BGP performance but did not study their interaction. In particular, others have shown that BGP update events occurring in a controlled fashion [31] and updates occurring in the wild [29] are strongly correlated with periods of high packet loss, increased path delay and loopy paths [44, 21, 43]. But none of this work studies the impact of such pathologies on various applications. Our study is the first to show that for real-time applications such as VoIP, BGP updates are correlated with a *significant* fraction of poor quality periods.

Additionally, VoIP has recently become an active research area, with existing work in three general categories: VoIP performance studies, studies of Skype's network, and other VoIP work. In contrast, this paper correlates VoIP quality with BGP updates. It also differs from prior work in one or more of the following: (1) It measures VoIP's performance across multiple domains in the Internet as opposed to a single backbone. (2) It reports user's perceived quality of voice, as opposed to network-related metrics like delay and jitter. (3) It is significantly larger than previous studies, with 50,000 calls between 24 locations, in US and Europe.

Most earlier work looks only at VoIP performance in a single backbone network. Boutremans et al. [10] examine VoIP's performance on the Sprint network. They find that due to over provisioning, congestion related packet-loss and latency are not significant issues for VoIP performance on a Tier-1 network. Markopoulou et al. [32] use data from 200 phone calls to study VoIP performance across various individual ISPs, and find that about 5% of their calls have inadequate quality. More recent work measures cross-domain

minutes. Many BGP events are very short and do not cause outages. The ones that cause outages, however, tend to last for a few minutes.

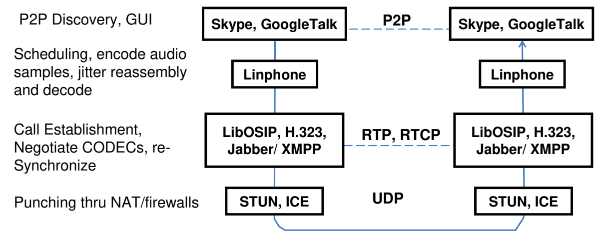

Figure 1: Components in a typical VoIP use scenario. For our experiments we used a customized Linphone client, running on top of libOSIP.

performance, but focuses only at packet level statistics such as delay and drops [5, 37, 39, 24, 33]. A few papers [30, 6, 37, 39, 25] show that, when possible, using multiple paths or dynamically switching paths based on real-time metrics improves VoIP performance.

Skype's popularity has motivated studies to understand its peer-to-peer VoIP network. The authors of [8] study Skype's login, firewall traversal, end-to-end encryption, codec, call establishment, and media transfer. They report that Skype's overlay is mainly used to by-pass NATs, and not to deal with performance problems. The authors of [11] propose a model to quantify user satisfaction by looking at call duration and evaluate their model on Skype traces. Others compare the performance of Skype with MSN VoIP clients [12]. They report that Skype preforms significantly better due to its codec and rate control mechanisms.

A few other papers have studied a variety of VoIP-related topics. Some have examined the performance of VoIP over wireless networks [26], its interaction with TCP and UDP [47], and how to provide the equivalent of "911" with a VoIP service [34].

However, none of the previous VoIP work has looked at the cause of availability problems for cross-domain VoIP calls, or, more specifically, their correlation with BGP updates.

3. Studying VoIP

In this section, we describe the components that VoIP software typically contains, the various algorithms for encoding and decoding voice traffic and the choices we made for our experiments.

3.1 VoIP Software

Figure 1 shows the four layers that make up a typical VoIP connection. At the lowest layer are mechanisms such as the ICE framework which uses protocols like STUN to punch holes through NATs and firewalls. Above this is a signaling protocol, typically SIP or H.323, that creates the RTP (Real Time Protocol)[38] connection, negotiates the codec to be used, and attempts to keep the sender and receiver synchronized using periodic RTCP messages. Just above the signaling protocol, sits a component such as Linphone [2] that handles the real-time scheduling. Typically, once every 20ms this component picks a sample from the sound input, encodes it using the pre-negotiated codec, and hands it off to be sent over UDP.

At the receiving end, packets arriving off the wire are reassembled in a jitter-concealing buffer, decoded and written

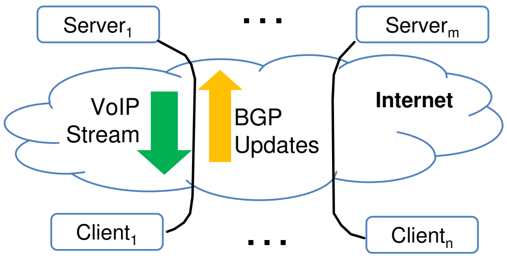

Figure 2: We pick nodes with BGP feeds to be the servers of VoIP streams. Recall that BGP updates flow in the reverse direction of the data path. This ensures that we collect BGP updates for the paths taken by the VoIP streams.

to the sound card, again once every 20ms. Finally, clients such as Skype and Google Talk handle all user interaction, peer-to-peer discovery of the caller/callee, and optionally route the call over an overlay.

Most known end-to-end VoIP clients are closed source, including Skype, MSN Messenger, Yahoo Messenger, AIM and Google Talk. Some, such as Skype, even implement proprietary protocols for all layers. Google, however, provides Libjingle, an open source "example" implementation of the Google Talk signaling and NAT hole-punching protocol, which uses the open sourced LinPhone [2] as its real-time component. Since all of our testbed machines have their own public IP addresses, we did not need the p2p discovery and NAT hole-punching functionality. Thus, in our experiments we used a modified Linphone client/server which uses the SIP signaling protocol. Our modified receivers dump the received VoIP stream to file instead of writing to the audio device.

3.2 Codecs

The choice of a VoIP codec trades off bandwidth for voice quality. Recent codecs claim to provide reasonable quality, while using as little as 6kbps. Table 1 summarizes the available options. Popular VoIP clients often support several of these codecs, and sometimes adaptively pick a codec based on path quality.

Our experiments use the G.711 (PCMU) codec, which is the highest quality codec in common use and uses about 64Kbps of bandwidth. We choose this codec because it is the default choice for most VoIP software packages and is the codec used by the PSTN. This choice produces conservative results in two ways. First, since the PCMU codec uses higher data rates than other codecs, it is likely to show better overall VoIP quality. Second, PCMU requires more bandwidth thus it is more sensitive to congestion than BGP, thus understating the relative effects of BGP updates.

Additionally, the PCMU codec specification advocates filling in lost frames (packets) with the previous frame. Linphone did not implement this feature, so we added it, after ensuring that it does improve voice quality.

4. Measuring VoIP Call Quality

The Mean Opinion Score (MOS) is the industry standard for measuring call quality [3]. It is expressed on a scale of 1 to 5, as follows.

CODEC	Data rate	MOS	Notes
G.711	64 Kb/s	4.1	Toll quality; Small coding delay
G.729	8 Kb/s	≈3.9	≈ 10ms coding delay
G.723	5.3/6.3 Kb/s	≈3.6-3.8	FEC; About 30ms coding delay
Speex	2.2-44.2 Kb/s	≈3.8	FEC; Variable Bit Rate for Low Bandwidth
iLBC	13.3/15.2 Kb/s	≈3.8	Upon packet loss, add FEC and increase rate

Table 1: Some popular VoIP codecs. VoIP clients typically support many of these codecs and some adaptively switch codecs.

MOS	Rating	Perceived Quality
4-5	Excellent	Toll Quality
3-4	Good	Cell Phone Quality
< 3	Fair	Unacceptable
< 2	Bad	Unintelligible

Ideally, the MOS is computed as the mean of the scores given by human judges who rate VoIP samples on a scale from 1 (bad) to 5 (excellent). Human studies however need controlled environments and are quite expensive to run. Instead, end-to-end tests of voice quality use standardized automated techniques[45], the most advanced of which is a signal processing based algorithm called PESQ (Perceptual Evaluation of Sound Quality) [23, 41]. PESQ compares the voice sample received over VoIP with the "original" voice sample to yield a MOS that has been shown to be highly correlated with MOS scores from human judges[23]. Thus, we use the PESQ algorithm for our experiments.

Finally, when the call is dropped, or the client cannot reconnect, there is no received voice stream to compute the PESQ score. We use a MOS of 0 for all such samples.

5. Experimental Setup

To measure the effect of naturally occurring BGP updates on VoIP, in as realistic a way as possible, over several weeks we ran over 50,000 one hour long phone calls, along approximately 100 paths between 24 RON hosts. We also collect the BGP updates affecting these paths, and correlate the performance of the VoIP streams with BGP updates.

(a) Measured Paths: We used the measurement sites from the RON [7] testbed shown in Table 2. We chose RON because BGP updates are available at several nodes. Also, the relatively light load on RON nodes allows the real-time VoIP clients to function with few CPU starvation incidents.

Further, it is known that Internet2 exhibits much better performance than the commercial Internet. RON contains many sites that belong to the commercial Internet, as shown in Table 2. To avoid data bias, we only include paths that do not have both ends connected to Internet2 [4], and thus traverse the commercial Internet.

(b) Call Configuration: We run a VoIP server on each of the five nodes with BGP feeds. The server is contacted by clients running on other nodes. We establish one SIP session over each of the paths available. For the practical

Name/Location	Upstream Provider(s)
Amsterdam, Netherlands	Global Crossing (GBLX)
Salt Lake City, Utah	xmission.net → Level3
Chicago, IL	GBLX
CMU, Pittsburgh, PA	AT&T, GBLX, Abilene ...
Laurel, MD	coloco.com → Cogent
Toronto, Canada	convoke.net → Cogent
Cornell, Ithaca, NY	Abilene ...
Delta, Canada	bigpipeinc.com → Level3
San Luis Obispo, CA	dwni.net → Cogent
Gatech, Atlanta, GA	Abilene ...
Austin, TX	hostway.com → Cogent
New York, NY	GBLX
London, UK	GBLX
MIT, Boston, MA	Genuity, Cogent, Comcast
San Jose, CA	megapath.net
Mount Vernon, IL	mvn.net
Hillsborough, NC	rr.com
NYU, New York, NY	Abilene ...
Tacoma, WA	opticfusion.net
New York, NY	speakeasy.net
UCSD, San Diego, CA	CENIC, Abilene ...
UMich, Ann Arbor, MI	Abilene ...
Univ. Utah, Salt Lake City, UT	Abilene ...
New York, NY	webair.net → GBLX

Table 2: We ran VoIP calls between 24 hosts. The five hosts in bold have BGP feeds and serve as the VoIP servers. Each VoIP call had at least one end-point with a BGP feed

reason of avoiding extremely long-running applications, we re-establish VoIP sessions every hour.

(c) Experiment Setup: Figure 2 shows our experimental setup. We run VoIP sessions along the paths for which we have BGP updates. We have BGP feeds at 5 of the RON nodes. BGP feeds tell us only about routing changes experienced by packets leaving the machine that has the feeds, and give us no information about the routing of packets coming to the machine. Thus, we locate the VoIP servers at the nodes with BGP feeds and run BGP clients on the other nodes. We make one sided phone calls with all voice data traveling from the servers, i.e., the machines with BGP feeds.

Since we do not have BGP feeds at the client nodes we do not capture the effect of BGP updates on the reverse path, i.e., the paths from client to server. This is acceptable since no voice data travels on the reverse path. Though the RTP protocol used by our VoIP client sends control packets in the reverse direction, such packets are infrequent and the protocol is more resilient to losses and/or high latency in this direction [38]. So, we believe that most routing problems impacting the connection occur on the forward path.

Further, for the purposes of this experiment, a 'BGP Update' on a path refers to any chatter received by the BGP daemon at the VoIP server's location for the destination prefix containing the VoIP client. We correlate the occurrence of such BGP updates with VoIP call quality on the path from the server to the client.

(d) Choosing Voice to Transmit: The quality of a voice sample depends on the uttered sentence as much as the network performance. To remove any bias caused by the choice of the utterance, all of our voice samples are utterances of the same sentence. The sentence we pick is a 3.5 second spoken sound sample of an English sentence (one utterance)

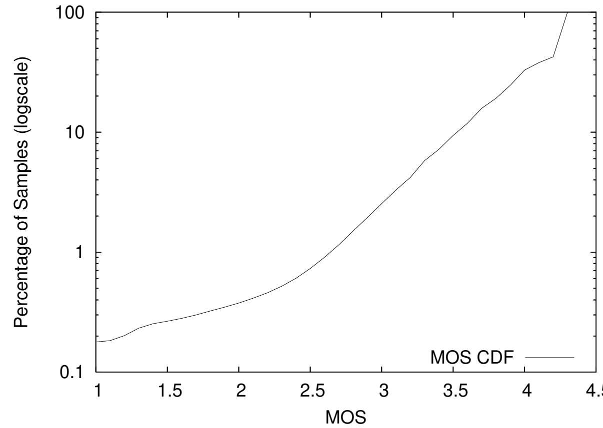

Figure 3: CDF of the MOS over all samples, shown with a log scale. It reveals that nearly 0.5% of samples have a MOS below 2, which reflects either an outage or unintelligible voice.

provided with the ITU reference PESQ implementation. To simulate a voice call, we modify the server to repeatedly replay this raw audio file.

(e) Eliminating CPU Scheduling Issues: We have observed instances in which CPU scheduling at the end-hosts introduces excessive jitter. For example, the server sending the packets is not scheduled for 40ms, or the client receiving the packets is not scheduled for 60ms. Since our focus here is on the impact of network events on VoIP performance, we have discarded samples that seem to have been affected by CPU scheduling issues.

6. Experimental Results

Our study supports two main conclusions. First, it shows that VoIP on the current Internet, though usable, does not perform well enough to replace the Public Switched Telephone Network (PSTN). Second, contradictory to expectations, VoIP is hindered as much by BGP as it is by network congestion. As a result, conventional QoS techniques alone, even if widely used, cannot eliminate current VoIP problems, and must, instead, be coupled with BGP-based solutions.

To simplify the discussion of our detailed results we first define the following terms.

VoIP Sample	3.5 second long period of received voice.
Outage	A contiguous period of voice samples with MOS $<$ 2, i.e., with unintelligible quality.
Silence Event	A contiguous period during which all packets are lost.
Call Abandonment	An instance of premature hang up caused by bad VoIP quality.
Near BGP Update	Within 10 minutes of a BGP update (also stated as *Correlated With a BGP Update*)

6.1 Call Quality & Availability

How good is VoIP? Our results show that voice quality in VoIP is acceptable, but availability is significantly lower

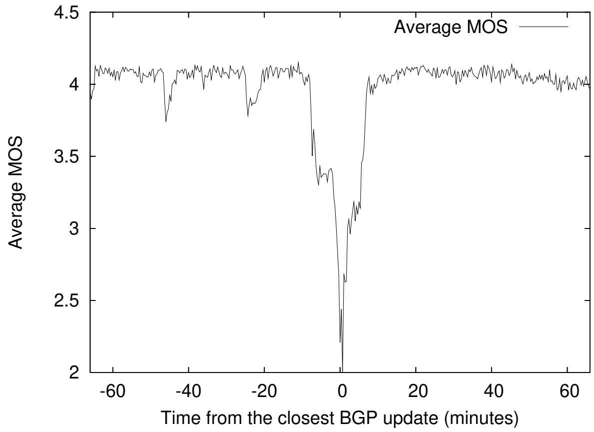

Figure 4: Average MOS across all samples as a function of the time between the sample and the closest BGP update. The figure shows that near a BGP update, the average MOS drops significantly. Both outages (MOS < 2) and bad quality samples (MOS < 3) occur around BGP updates.

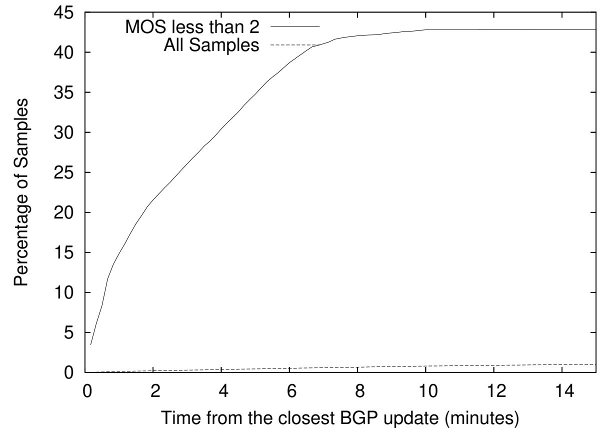

Figure 5: CDF of the time from closest BGP update (in minutes), taken over all outage samples. It shows that almost 50% of outage samples are within 10 minutes of a BGP update. In contrast, less than 1% of all samples are within 10 minutes of a BGP update.

than the PSTN. Figure 3 shows that 97% of VoIP samples have a quality similar to or better than cell phones—i.e., a MOS higher than 3. On the flip side, VoIP availability is relatively low. The same figure shows that about 0.5% of samples have a MOS lower than 2, which implies an unintelligible utterance, and hence an outage. This means that users experience on average 5-10 minutes a day of outage. Though such a level of availability is acceptable for a free casual-use service, it is at least one order of magnitude lower than the availability of PSTN land-lines [36].

6.1.1 Correlation of Quality and BGP Updates

If drops in VoIP quality are highly correlated with BGP updates, the experiment should reveal two things. First, the data should show that, when a BGP update occurs, on average there is a significant drop in MOS. Additionally, the data should also show the converse, that when there is a significant drop in MOS, it is likely that a BGP update is nearby.

(1) BGP Update ⇒ Low MOS: Figure 4 shows that the average MOS near a BGP update drops from over 4.2 for samples far from a BGP update, to about 2 for samples near a BGP update. Thus, for samples far from a BGP update, the average MOS is toll quality, while near a BGP update the average MOS drops to unintelligible.

(2) Low MOS ⇒ Nearby BGP Update: While the results in Figure 4 are somewhat expected, Figure 5 shows a surprising result; almost 50% of the periods with outages (i.e., a MOS < 2) occur within 10 minutes of a BGP update. One may wonder whether BGP updates occur frequently enough on our paths such that it is expected for any given sample to be within 10 minutes of a BGP update independent of whether the sample is good or bad. Figure 5 shows that such intuition is wrong; less than 1% of our voice samples are within 10 minutes of a BGP update. Furthermore, the average interarrival time between updates on the measured paths is about 18 hours. This emphasizes the striking result that while BGP updates are widely spread, almost half the outage samples are concentrated around them.

6.1.2 Explaining BGP's Significant Quality Effect

Since BGP updates only occur about once a day on each of the paths that we studied, to those unfamiliar with BGP it may seem surprising that such a significant portion of outages can be correlated with BGP updates. This stems in part from the ability of the VoIP client to conceal packet-loss resulting from non-BGP sources, in particular congestion, as discussed in §6.2. More importantly however, BGP's dynamics allow path change events to cause significant periods of packet-loss while BGP convergences, even when the underlying network is completely connected [44].

Much prior work shows empirical evidence that BGP convergence can cause loops, delays, and disconnectivity for tens of minutes [29, 44, 16, 20]. This is because the path-vector policy-driven nature of BGP hinders an autonomous system (AS) from quickly eliminating alternate paths that are no longer available. Instead, BGP convergence often goes through an extended period of moving to and announcing alternate paths that are later found to be not available. To stop the ASes from quick announcements that are more likely to be spurious, BGP uses a timer, call the MRAI (Minimum Route Advertise Interval), that forces ASes to wait longer and collect more information before re-announcing a path to a neighbor. While the MRAI timer does succeed in suppressing spurious announcements, a large value for the timer delays overall convergence. To make matters worse, prior work [31] shows that choosing an optimal MRAI value to converge as quickly as possible is tricky. In fact, the optimal MRAI is network specific, and no single value of MRAI would be optimal for all networks [19].

6.1.3 Non-BGP Reasons For Bad Quality

Given that approximately 50% of outage samples are correlated with BGP updates, it is natural to ask what is causing the other 50%. While it is difficult to determine the cause of these outages, it is likely the majority of them are caused by network congestion. Also, some may be due to slow intra-domain routing convergence or intra-domain network failures.

One interesting trend, however, is worth noting. Unintelligible samples (i.e., outages) contain many packets that

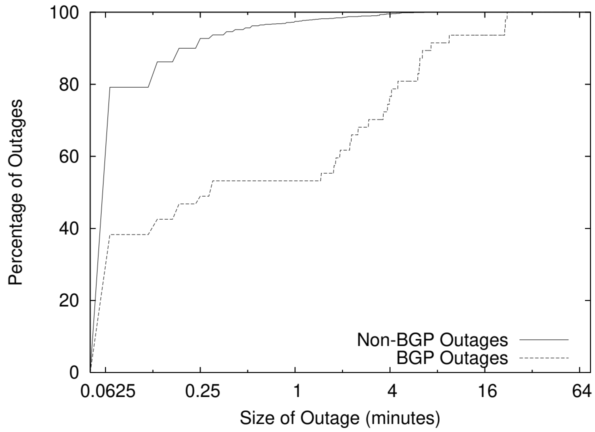

Figure 6: CDF of outage length (in minutes) for samples within 10 min. of a BGP update, and samples greater than 10 min. from a BGP update. It shows that outages near a BGP update last longer than those uncorrelated with a BGP update.

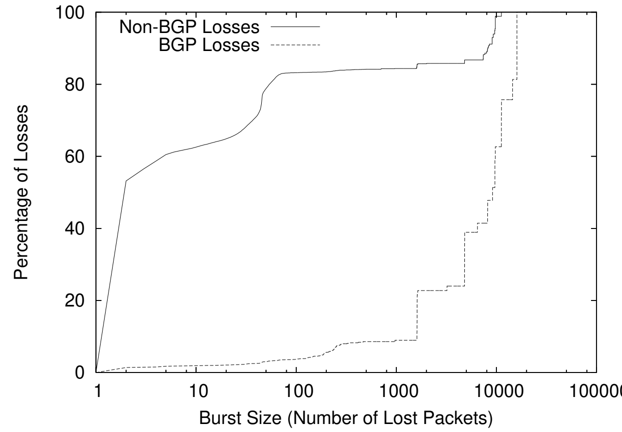

Figure 7: CDF of loss burst length (in packets) for losses within 10 min. of a BGP update, and those more than 10 min. from a BGP update. It shows that at the packet level, losses correlated with BGP burst much longer than those uncorrelated with BGP.

could not be played either because they were lost or they arrived too late. In outage samples not correlated with a BGP update, only 15% of the *unplayable* packets were actually lost, while almost 85% were unplayable simply because the packet arrived too late. In contrast, almost 70% of the unplayable packets in the BGP-correlated outage samples were lost. More intelligent jitter buffer management might allow many of the delayed packets to be played, avoiding many of the non-BGP outages.[2] But little can be done to allow a lost packet to be played. Thus, we believe the results in this section to represent a conservative estimate of the effect of BGP on VoIP quality and reliability.

6.2 Duration of Bad Quality

As one might expect, when problems occur near a BGP update, they tend to last far longer than those uncorrelated with a BGP update. This simple property however, underscores two important points, one at the level of VoIP samples, and one at the level of individual packets.

At the VoIP sample level, the increased length of outages near BGP updates highlights user's frustration caused by BGP updates. In particular, Figure 6 shows that over 90% of non-BGP outages last less than 10 seconds. In contrast, most BGP outages last over a minute and about 30% of BGP outages persist for greater than 4 minutes. These durations translate directly to the callers' ability to either call back immediately after an outage, or to "wait it out", hoping the quality will get better. While it may be reasonable to wait out a 10 second outage, callers cannot be expected to wait out a 4 minute outage, thus illustrating the increased frustration BGP correlated outages cause vs. non-BGP correlated outages.

At the individual packet level, packet loss periods[3] near BGP updates are also much longer. This helps to explain how almost 50% of outage samples correlate with BGP updates even though only 15% of lost packets correlate with BGP updates. Recall that an outage sample– meaning a 3.5s voice sample with a MOS score less than 2– generally occurs only when many packets within the sample are lost. Figure 7 shows that more than 50% of the packets lost away from a BGP update are individual packet losses, and more than 60% of them are part of a burst of 3 packets or less. Such bursts are short enough that that they can be concealed using automated techniques such as Forward Error Correction (FEC) or other loss concealment techniques [27]. In contrast, 90% of the lost packets near BGP updates are part of a burst of 1500 packets or longer, translating into at least 30 seconds of continuous loss.

This explains why the relatively randomly distributed packet-loss caused by congestion usually does not affect the VoIP transmission enough to create outage samples, while the BGP correlated packet-loss is bursty enough that most samples that see BGP correlated loss, end up unintelligible. In particular, studies show that for random packet-loss, when linear interpolation is used to replace lost packets, more than 30% packet-loss can be tolerated while still maintaining MOS scores greater than 2.4 [27]. Advanced techniques such as forward error correction (FEC) can do even better. In contrast, when loss bursts last longer than the jitter buffer size, which is typically on the order of 3-6 packets or 60-120 ms, loss concealment and FEC techniques cannot help[40] .

6.3 Call Abandonment

Another important metric for VoIP performance is the ability to conduct long conversations without being forced to abandon the call. Poor VoIP quality leads callers to prematurely hang up. Since we do not have actual human users, we could not measure hang up rates. Instead, we analyze this effect using the telephony industry's standard for estimating hang up rates. ITU standard E.855 [22] estimates that the likelihood a user does not hang up the call is exponentially distributed with d, the length, in seconds, of experienced silence events:

$$P[\text{hold on}] = e^{-d/17.26}. \tag{1}$$

[2] Linphone's jitter management techniques are pretty basic. Playout is delayed by three samples (or 60ms) under normal conditions. Further, the playout delay is adaptively increased only if the path has high average jitter [2].

[3] Periods during which all packets are lost.

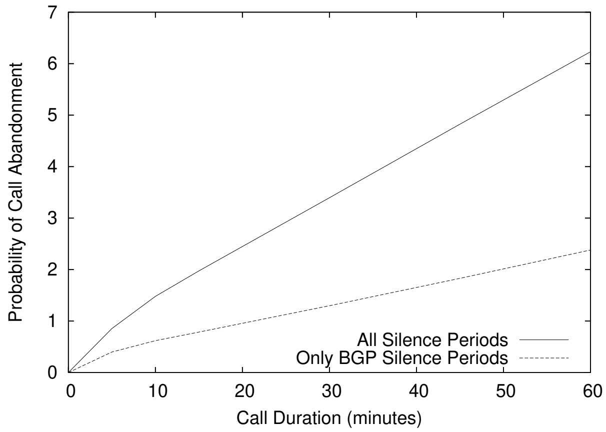

Figure 8: Percentage of calls abandoned as a function of call length for all silence periods, and only silence periods within 10min of a BGP update. Percentages are computed using Eq. 2 and averaged across all possible call start times on all paths. The figure shows that 6.2% of one-hour calls are abandoned.

We combine this estimate with our voice data to estimate the likelihood that a user can conduct a conversation of duration x without hanging up. A call of duration x beginning at time t will not be abandoned only if the user does not hang-up during any of the silence events that happen between t and $t + x$. Let $d_1, d_2, \ldots$ be the silence events that occur between t and $t+x$, then the probability of not hanging up during that period $P_x(t)$ is:

$$P_x(t) = \prod_i P(\text{hold on during } d_i) = e^{-\frac{\sum_i d_i}{17.26}}. \qquad (2)$$

We then average over all possible starting times t to estimate P_x—the probability that a call of duration x goes through.

Computing the probability of call abandonment as a function of the call duration shows that in 6.2% of one-hour calls (1 out of 16), the caller will prematurely hang up because of silence events. More generally, Figure 8 shows the probability of call abandonment, $1 - P_x$, as a function of call duration, computed using Eq. 2. From this, we can see that a little less than half of the call abandonment is correlated with BGP updates, consistent with the overall number of outages.

Finally, we note that our estimate of the probability of call abandonment is conservative. Silence events are not the only the reason for abandoning a call. A user may hang up because of bad quality. Since there are no models for call abandonment as a function of VoIP quality, one would need human subjects to estimate such an effect.

7. Concluding Remarks

We have shown that VoIP problems correlated with BGP updates occur often enough, cause enough calls to be abandoned, and create long enough unavailability that we believe it is difficult to replace a significant fraction of land-line phone service with cross-domain VoIP without fixing the performance problems correlated with BGP.

Though our study of VoIP performance has limitations, we believe our results provide conservative estimates. In particular, we stream voice only in one direction though two-way voice is the norm in telephony. Fully duplex VoIP is likely to demand more stringent performance; for example users will hang-up a call when a problem happens on either the forward or the reverse paths. Second, RON nodes tend to have higher bandwidth and more stable BGP routing, making their performance better than average Internet paths. We tried to limit this factor by using only non-Internet2 paths. Still our data may show slightly better performance than average Internet paths. Finally, newer codecs such as iLBC and Speex, employ sophisticated forward error correction techniques that can recover from random packet losses; this has the potential to improve overall performance of VoIP but may also make it more crucial to avoid the bursty losses caused by BGP events.

Our results show a strong correlation between VoIP intelligibility and BGP updates, motivating the need for a solution. But, how do we address quality degradation due to BGP? A potential solution may use Inter-domain overlays [42, 8]. It is unclear, however, whether overlays can react soon enough –50% of callers hang up after 12 seconds of outage [22]– and without excessive probing overhead. Note that Skype's overlay is mainly used to by-pass NATs, as opposed to performance problems [8]. Additionally, it seems that the VoIP market may increasingly be dominated by the ISPs [15], and traditionally ISPs have been unable to build inter-domain overlays, which require them to cooperate tightly [13]. The alternate option is to modify BGP to eliminate transient disconnectivity and ensure fast recovery. The ISPs' desire to capture the VoIP market may provide the necessary incentive for them to fix BGP. We believe that inter-domain multipath routing [46] and our recent work on pre-computed BGP fail-over paths [28] provide significant steps toward that goal.

8. Acknowledgments

We thank Hariharan Rahul and Mike Walfish for their comments on the paper. Kushman and Katabi are supported by the NSF award CNS-0448287. Kandula is supported by a gift from Cisco Systems. The opinions and findings in this paper are those of the authors and do not necessarily reflect the views of NSF or Cisco.

9. References

[1] Cable Operators Embrace VoIP. http://www.technewsworld.com/story/40374.html.

[2] Linphone: Telephony on Linux. http://www.linphone.org.

[3] Mean Opinion Score. http://en.wikipedia.org/wiki/Mean_Opinion_Score.

[4] Abilene. http://monon.uits.iupui.edu/.

[5] Y. Amir, C. Danilov, S. Goose, D. Hedqvist, and A. Terzis. An Overlay Architecture for High Quality VoIP Streams. In *IEEE Transactions on Multimedia*, December 2006.

[6] Y. Amir et al. 1-800-OVERLAYS: Using Overlay Networks to Improve VoIP Quality.

[7] D. Andersen, H. Balakrishnan, M. F. Kaashoek, and R. Morris. Resilient Overlay Networks. In *SOSP*, 2001.

[8] S. Baset and H. Schulzrinne. Analysis of the Skype Peer to Peer Internet Telephony Protocol. In *INFOCOM*, 2006.

[9] O. Bonaventure, C. Filsfils, and P. Francois. Achieving sub-50ms Recovery upon BGP Peering Link Failures. In *Co-Next*, 2005.

[10] C. Boutremans, G. Iannaccone, and C. Diot. Impact of link failures on VoIP performance. In *NOSSDAV*, 2002.

[11] K.-T. Chen et al. Quantifying Skype User Satisfaction. In *SIGCOMM*, August 2006.

[12] W.-H. Chiang et al. A Performance Study of VoIP Applications: MSN vs. Skype. In *MULTICOMM*, 2006.

[13] D. Clark, W. Lehr, P. Faratin, S. Bauer, and J. Wroclawski. The Growth of Internet Overlay Networks: Implications for Architecture, Industry Structure and Policy. In *TPRC*, 2005.

[14] P. Drew and C. Gallon. Next-Generation VoIP Network Architecture. Technical Report MSF-TR-ARCH-001-FINAL, Multiservice Switching Forum, March 2003.

[15] T. Espiner. Skype's Market Share Halves, February 2006. http://news.zdnet.co.uk/communications/broadband/0,39020342,39250116,00.htm.

[16] A. Freedman. Edge/Core Update Propagation, Churn vs. Performance, 2001. Presentation at Internet Statistics and Metrics Analysis workshop (ISMA).

[17] Frost and Sullivan. North American Wholesale Long Distance Voice Service Market - Gaining a Competitive Advantage Through Migration to VoIP, Dec 2004.

[18] Frost and Sullivan. Trends in Wireline Substitution - North American Markets, August 2005.

[19] T. Griffin and B. Premore. An experimental analysis of bgp convergence time. In *Proceedings of IEEE ICNP*, 2001.

[20] U. Hengartner, S. Moon, R. Mortier, and C. Diot. Detection and Analysis of Routing Loops in Packet Traces. Technical Report TR02-ATL051001, Sprint ATL, 2002.

[21] P. Huang, A. Feldmann, and W. Willinger. A Non-intrusive, Wavelet-based Approach to Detecting Network Performance Problems. In *IMW*, 2001.

[22] International Telecommunication Union. Connection Integrity Objective for International Telephone Service. E.855, 1988.

[23] International Telecommunication Union. Perceptual Evaluation of Speech Quality. In *ITU-T Rec. P.862*, 2001.

[24] W. Jiang and H. Schulzrinne. Assessment of VoIP Service Availability in the Current Internet. In *PAM*, 2003.

[25] M. Karol, P. Krishnan, and J. Li. Using Overlay Networks to Improve VoIP Reliability. In *LNCS*, 2004.

[26] T. Kawata, S. Shin, A. G. Forte, and H. Schulzrinne. Improving the Capacity for VoIP Traffic in IEEE 802.11 Networks with Dynamic PCF. In *IEEE WCNC*, 2005.

[27] K. Kondo and K. Nakagawa. A Speech Packet Loss Concealment Method Using Linear Prediction. In *IEICI Trans. Inf. and Syst.*, volume E89-D.

[28] N. Kushman, S. Kandula, D. Katabi, and B. Maggs. R-BGP: Staying Connected in a Connected World. In *NSDI*, April 2007.

[29] C. Labovitz, A. Ahuja, A. Bose, and F. Jahanian. Delayed Internet Routing Convergence. In *SIGCOMM*, pages 175–187, 2000.

[30] Y. Liang, E. Steinbach, and B. Girod. Multi-stream voice over IP using Packet Path Diversity. In *IEEE Multimedia Signal Processing*, 2001.

[31] Z. Mao, R. Govindan, G. Varghese, and R. Katz. Route-flap Damping exacerbates Internet Routing Convergence. In *ACM SIGCOMM*, 2002.

[32] A. Markopoulou, F. Tobagi, and M. Karam. Assessment of VoIP Quality over Internet Backbones. In *INFOCOM*, 2002.

[33] I. Marsh and F. Li. Wide Area Measurements of Voice over IP Quality. Technical report, Swedish Inst of Computer Science, 2003.

[34] M. Mintz-Habib, A. S. Rawat, H. Schulzrinne, and X. Wu. A VoIP Emergency Services Architecture and Prototype. In *ICCCN*, 2005.

[35] E. Osborne and A. Simha. *Traffic Engineering with MPLS*. Cisco Press, 2002.

[36] PacketCable. VoIP Availability and Reliability Model for PacketCable Architecture. Technical Report PKT-TR-VoIPPAR-W05-001012, 2000.

[37] R. K. Rajendran et al. Performance Optimization of VoIP using an Overlay Network. Technical report, NEC, 2005.

[38] H. Schulzrinne et al. RTP: A Transport Protocol for Real-Time Applications. In *IETF RFC 1889*, 1996.

[39] S. Tao et al. Improving VoIP Quality Through Path Switching. In *INFOCOM*, 2005.

[40] Teck-Kuen and D. C. Pheanis. Effects of Loss Characteristics on Loss-Recovery Techniques for VoIP. In *ICNICONSMCL*, 2006.

[41] M. Varela, I. Marsh, and B. Grnvall. A Systematic Study of PESQ's Performance. In *MESAQIN*, 2006.

[42] Vonage – IP Telephony/Voice over IP. http://www.cisco.com/en/US/tech/tk652/tk701/technologies_case_study09186a00800b559e.shtml.

[43] F. Wang, N. Feamster, and L. Gao. Quantifying the Effects of Routing Dynamics on end-to-end Internet Path Failures. Technical Report TR-05-CSE-03, University of Massachusetts, 2006.

[44] F. Wang, Z. M. Mao, J. W. L. Gao, and R. Bush. A Measurement Study on the Impact of Routing Events on End-to-End Internet Path Performance. In *SIGCOMM*, 2006.

[45] A. Wittman. Assuring VoIP Quality: Not There Yet. IT Architect, March 2005.

[46] W. Xu and J. Rexford. Multi-path Interdomain Routing. In *SIGCOMM*, 2006.

[47] X. Zhang and H. Schulzrinne. Voice over TCP and UDP. Technical Report CUCS-033-04, Columbia University, 2004.