Assessment of VoIP Service Availability in the Current Internet

Wenyu Jiang
Department of Computer Science
Columbia University
Email: wenyu@cs.columbia.edu

Henning Schulzrinne
Department of Computer Science
Columbia University
Email: hgs@cs.columbia.edu

***Abstract*— We evaluate the availability of voice over IP (VoIP) service typically achieved in the current Internet. Service availability is examined using several metrics, including call success probability, overall packet loss probability, the proportion of time the network is suitable for VoIP service, and call abortion probability induced by network outages. Our major findings are: first, packet losses are not rare events, and it is generally worse on international paths. Secondly, network outages make up a non-negligible portion of packet losses. While most outages are short, some are extremely long and make up the majority time of all outages. This implies when a service becomes unavailable, the Mean Time To Restore (MTTR) can sometimes be very high. About one third of the outages happen in symmetry, and outages tend to occur at the edges rather than in the middle of the network. Thirdly, although research networks such as Internet2 has much lower delay and loss than the public Internet, the effect of network outages on both types of networks is almost the same. Finally, we will show that when considering calls aborted due to network outages, the overall service availability drops by a significant margin, from about 99.5% to 98%.**

I. Introduction to Voice over IP and Service Availability

In recent years, voice over IP (VoIP) has gained increasing popularity. But the public switched telephone network (PSTN) has established a strong impression to the general public in terms of its high availability and reliability despite its age. If VoIP were to successfully replace the PSTN, it has to meet several stringent requirements, in particular high service availability. Although there are numerous literature studying Quality of Service (QoS) in the Internet [10], [9], [3], little has been done examining the aspect of service availability. When transitioning from a technical hobby to a regular service, VoIP has to provide a high degree of availability for day-to-day business needs.

A. Definitions

Broadly defined, availability is the proportion of time that a service is available for use. Reliability, by comparison, measures how long a service can stay up before it is disrupted. Therefore, reliability is measured in terms of Mean Time Between Failures (MTBF) and Mean Time To Restore (MTTR). We can then define availability as follows [1]:

$$Availability = \frac{MTBF}{MTBF + MTTR} \tag{1}$$

In the context of telephony, availability is the probability that a call can be established successfully on first attempt, excluding user factors like callee busy or no pick-up. Therefore, we may use the following alternative definition:

$$Availability = \frac{\text{\# of successful calls}}{\text{\# of first call attempts}} \tag{2}$$

Although Eq. 2 appears different from Eq. 1, the two definitions are effectively the same over the long run. This is because in Eq. 2, we are in essence sampling the service for its “up” and “down” status.

Among the telecommunications equipment vendors, the five 9’s (99.999%) availability has been advertised for decades. Five 9’s implies a total downtime of only 5 minutes and 15 seconds per year. A laudable goal it is, but is it easily achievable? In contrast to what many people may think, the five 9’s really means the availability of the local switching equipment, such as a Private Branch eXchange (PBX) or local central office (CO) switch. Across an entire telephone network, such as a U.S. domestic telephone network, it is difficult to achieve anywhere near five 9’s, because there are many network elements that may break and result in call failure. Assuming the components in a call chain are independent in terms of availability, we can define the end-to-end (handset-to-handset) availability as follows:

$$A_{e2e} = A_{h1} \cdot A_{local1} \cdot A_{network} \cdot A_{local2} \cdot A_{h2} \tag{3}$$

where A_{h1} and A_{h2} are availability figures of the caller and callee telephone handsets, respectively. For regular (analog) telephones, their values generally should be 1, but for IP phones, the values would depend primarily on the stability of its software, which is a software engineering problem and beyond the scope of this paper. A_{local1} and A_{local2} are availability figures of caller’s and callee’s local PBX or CO switch, and this is where the five 9’s should be expected. In the case of VoIP, this component corresponds to a dedicated call server such as a SIP [13] proxy server, and it may be implemented in hardware in the form of an IP PBX. Because VoIP is protocol and software intensive, the reliability of this component would again depend mostly on its code stability and therefore not the focus of this paper. Finally, $A_{network}$ is the availability of the network, whether it is the PSTN or an IP network such as the Internet. The evaluation of this component on the Internet is the central objective of this study.

B. Typical Performance

There is no commonly agreed-upon value for telephone network availability ($A_{network}$), but we can get a sense from what is offered in the real world. AT&T, for example, advertises 99.98% availability in 1997 on its US domestic telephone network [1]. That is, when a user dials a domestic number, there is a 99.98% chance it will succeed on the first attempt. When evaluating VoIP availability, Audin [1] uses a value of 99.9% (IP frame relay), based on service level agreements posted by certain providers [2]. So the US domestic network availability appears to be on the order of three to four 9's. Consequently, the "bar" of availability for VoIP over the Internet need not (and probably cannot) be at five 9's, but one to two orders lower.

For international calls, it is much more difficult to set a pre-determined availability level, as it would depend strongly on how good the PSTN (or IP) network is in both the caller and callee countries. The World Bank Group provides some relevant statistics, for example, local call completion ratio (CCR) for certain countries [3]. These values provides an upper bound on each country's telephone service availability, since it does not consider domestic long-distance calls. As an example, in 1992 the local CCR for United Kingdom is 99.8%, and for France it is 99.4%, whereas it is only 61.0% for South Africa. CCR may also be defined as the ratio of calls answered (picked up) vs. call attempted. In such case, the local CCR would be much lower, usually 60-75% even on a good network, as the callee may not be around or may be busy on the phone. However, we do not define availability or CCR in this way, because in our measurements the calls are generated and answered automatically by agents running on the measurement nodes.

VoIP is sometimes likened to mobile telephony, because they both possess some advantage over the fixed line telephone network. VoIP has an edge in cost savings, while the latter provides mobility and flexibility. Therefore it is inspiring to learn about the availability achieved on mobile telephone networks. In a 2002 survey of mobile network operators in United Kingdom [4], the availability (defined there as rate of successful call setups) ranges from 97.1% to 98.8% on the national level, depending on the provider.

It should be noted that the overall availability sometimes does not describe the end-user experience well. For instance, users may be more upset with longer outages, when the total duration of outages is the same. Moreover, if a call is interrupted by a network outage (which we define here as long, consecutive loss bursts), the user will eventually abort the call if the interruption lasts too long. We will discuss these issues in the subsequent sections.

[1] http://www.att.com/network/standrd.html#np

[2] https://www.sprintbmo.com/bizpark/page/general.jsp?general_id=1017

[3] http://www.worldbank.org/html/opr/pmi/telecom/teleco07.html

[4] http://www.oftel.gov.uk/publications/research/2002/call_survey/

C. The Importance of Availability

Service availability is an important concept in telephony for several reasons. First, the telephone service has become a vital and integral component of our daily life, upon which businesses and individuals depend. Another reason is that in most situations, something is better than nothing. A call with low quality is preferable to not being able to make a call. In addition, the most important factor causing low quality, namely packet loss, can be ameliorated using quality improvement techniques such as FEC [8]. Service unavailability, by contrast, represents a bigger problem. Assuming the local switching equipment is reliable, it means the network is unavailable or inaccessible. This usually implies long outages spanning tens to thousands of packets, not isolated losses or short loss bursts. For FEC to recover lost packets efficiently, the number of its redundant packets in each block has to be larger than the typical loss burst length. Therefore FEC will not work during long outages due to the excessive delay it would introduce. From a practical point of view, for outages beyond a few seconds, FEC is no better than retransmissions.

Given the importance of service availability in telephony, if VoIP were to successfully replace or at least compete with the PSTN, it has to match the bar of service availability. Yet there is a lack of study on this topic. Therefore we have performed a series of Internet measurements to assess the level of availability that the current Internet can provide for VoIP. We examine several performance metrics, including call success probability, overall packet loss probability, the proportion of time the network is suitable for VoIP service, and call abortion probability induced by network outages.

For the remainder of this paper, we will describe in Section II the setup of our Internet measurement, and in Section III the metrics we use to evaluate VoIP service availability along with the measurement results. Section IV lists related work in Internet measurements. Finally, we conclude the paper in Section V.

II. VoIP Availability Measurement Setup

To evaluate the service availability and quality of VoIP, we obtained a total of 14 measurement nodes (Unix machines and PCs), listed in in Table I. Due to quadratic nature of a mesh, we can measure $O(N^2)$ one-way paths, although it should be noted that some nodes are PCs subject to power-off by their owners. These nodes also provide a good mix of both research and commercial (public) parts of the Internet, allowing us to compare the characteristics of the two types of networks. Although not all shown in Table I, the access link bandwidths of all research network nodes are much larger than residential broad-band connections such as Cable modem and ADSL.

We developed active measurement software agents to automatically make simulated voice calls between these endpoints. The call duration was 3 minutes during the initial phase of measurements, and soon changed to 7 minutes. Both values are within reasonable range for typical long-distance calls [6]. Initially, the agents make calls every 5 minutes with each call

Node name	Location	Connectivity	Network
columbia	Columbia Univ, NY	$\geqslant$ OC-3	I2
wustl	Washington Univ, St. Louis		I2
unm	Univ of New Mexico		I2
epfl	EPFL, Switzeland		I2+
hut	Helsinki Univ of Technology		I2+
rr	NYC	Cable modem	Commercial
rrqueens	Queens, NY	Cable modem	Commercial
njcable	Green Brook, New Jersey	Cable modem	Commercial
newport	Newport, New Jersey	ADSL	Commercial
sanjose	Sanjose, California	Cable modem	Commercial
suna	Kitakyushu, Japan	3Mb/s access link	Commercial
sh	Shanghai, China	Cable modem	Commercial
Shanghaihome	Shanghai, China	Cable modem	Commercial
Shanghaioffice	Shanghai, China	ADSL	Commercial

TABLE I

LIST OF MEASUREMENT NODES, I2+ IS INTERNET2 PLUS ITS PEERING NETWORKS

lasting 3 minutes, leaving 2 minutes for transfering trace files to a central file server. Later, the call interval is changed to 10 minutes, giving 3 minutes for trace file transference. Therefore the main reason for switching to 7 minutes is to measure the network more continuously. However, we did not choose a very long call duration or call interval (e.g., an hour), because some of our test nodes are PCs that may be turned off by their owners, and we do not want to create too many "interrupted" calls.

We choose the packet interval to be 40 ms, as that is small enough to capture short term delay/loss characteristics while large enough to keep the size of the measurement data small. The bit rate at UDP layer for each call is 10.8 kb/s, which reflects the use of a low bit-rate codec. We also vary the payload size every other packet, to examine the effect of packet size on path behavior. This is useful in verifying the effectiveness of forward error correction (FEC) [12], since FEC introduces some bandwidth overhead to the original packets.

The calls are scheduled to start simultaneously for each batch of tests, so that we can verify if $A \to B$ experiences a problem, whether it is reflected in $A \to C$ as well. The software also randomizes the destination it calls and limits the maximum number of calls for a node at any time, to avoid congestion of its access link. Therefore, our measurement is full-mesh only in a logical sense, and is more scalable.

We collected over two months of data, from September 10, 2002 until December 6, 2002, totaling over 13,500 call hours of data. Therefore, it provides enough samples to give a reliable picture of the part of the Internet being measured.

III. Measurement Results

A. Call Success Probability

As discussed in Section I-A, the five 9's refers to local switch (PBX or CO) availability, whereas the network availability is usually lower, at three to four 9's on the U.S. domestic telephone network. Network availability is usually measured by call success probability on first call attempt, as per Eq. 2. Therefore, for each simulated voice call, our measurement software (caller) sends a call request to the callee. If it sees no response, either there is a network failure, or the callee machine is powered off by its owner. Later, each unresponded call attempt is checked offline against the callee's liveness log. If the callee software was indeed alive (running) at the time, then it can be safely assumed that the reason was network failure. To prevent occasional packet losses from being interpreted as network failure, the call requests are transmitted up to three times with one second timeouts, making call failure under occasional packet losses highly unlikely.

During the entire measurement period, 62,027 calls were successfully made, and there were 292 unsuccessful calls due to network failure. Therefore, at a very high level, we can estimate the overall service availability as 99.53%. Clearly, it is far from being practical to ask the Internet to deliver five 9's availability as found on PBX equipment.

When considering different types of paths, such as Internet2 only, Internet2+ (including those peered with I2), and the commercial Internet, very interestingly, the call success probability remains roughly constant.

Table II lists the detailed numbers. Note that a path between a research network node and commercial network node is a commercial path. The value for US domestic commercial case (99.39%) renders itself an applicable comparison to the 99.98% value advertised by AT&T. But apparently, the Internet still has some way to go before it can match a well-engineered PSTN. If we "lower" the bar a little and compare it to the mobile phone, whose availability in UK ranges form 97% to 99% (Section I-B), the Internet is actually slightly better. Hence the morale is, if most users treat VoIP on par with mobile telephony, it would be considered fairly good. But it is obviously not the aim of this paper to speculate on the validity of this claim.

Network/path type	Call success probability
All	$\frac{62027}{62027+292}$ = 99.53%
Internet2	$\frac{8505}{8505+41}$ = 99.52%
Internet2+	$\frac{27320}{27320+121}$ = 99.56%
Commercial	$\frac{34699}{34699+171}$ = 99.51%
Domestic (US)	$\frac{19858}{19858+110}$ = 99.45%
International	$\frac{42165}{42165+175}$ = 99.58%
Domestic commercial	$\frac{11353}{11353+69}$ = 99.39%
International commercial	$\frac{23345}{23345+95}$ = 99.59%

TABLE II

CALL SUCCESS PROBABILITY ON FIRST CALL ATTEMPT WITH RESPECT TO NETWORK/PATH TYPE

In Table II, the US domestic commercial case has the lowest call success probability, but the difference from the overall mean (99.53%) is small, only 0.14%. It is also noted that research networks (Internet2 and Internet2+) does not have significantly higher availability than commercial networks. This suggests that the cause for call failure (unavailability), which is most likely due to long outages, are roughly the same on both types of networks.

B. Overall Network Loss

The difference between the PSTN and the IP network is that once a call is established, a PSTN connection generally guarantees very good quality for the duration of the call. The Internet, by contrast, provides only best effort service, so packets are subject to loss and delay, thereby reducing quality. In our measurement, the overall network packet loss probability across all calls is 0.56%, or 99.44% delivery probability. Although this is not great compared to the 99.9% figure used by Audin [1] for frame relay, it is not too high either. For example, a different ISP [5] quotes $\leqslant$ 0.7% long term loss probability. Either way, packet losses are not rare events in the Internet. It is also interesting to observe that from Sec. III-A, the initial call failure probability is 1-99.53% = 0.47%, which is quite close to the 0.56% loss probability for established calls. The overall loss rates are listed for each type of network path in Table V, for ease of comparison with network outages, to be discussed in Section III-C. In Table V, Internet2 is the unarguable winner with only 0.227% overall loss, whereas domestic and international commercial part of Internet perform among the worst.

However, a single, average value is often misleading. In practice, many calls experienced quality that may be bad enough to be considered service interruptions. To define it quantitatively, let us examine the distribution of loss probability when it is computed every 10 seconds. Table III shows proportion of time when the network loss probability is below a certain level. For example, 97.48% of the time the network has $\leqslant$ 5% loss. When using a low bit-rate codec, which is a necessity when congestion occurs, 5% loss can cause a noticeable degradation [8] in quality, possibly enough for the user to drop the call. Therefore, if we assume a 5% tolerance level for the user, then availability drops to 97.48%. At 10% tolerance, availability becomes 99.16%.

[5]http://www.att.com/network/#reliability

Network/path type	%time loss is below			
	0%	5%	10%	20%
All	82.3%	97.48%	99.16%	99.75%
Commercial	78.6%	96.72%	99.04%	99.74%
Internet2	97.7%	99.67%	99.77%	99.79%
Internet2+	86.8%	98.41%	99.32%	99.76%
Domestic (US)	83.6%	96.95%	99.27%	99.79%
International	81.7%	97.74%	99.11%	99.73%
Domestic commercial	73.6%	95.03%	98.92%	99.79%
International commercial	81.2%	97.60%	99.10%	99.71%

TABLE III

DEPICTING SERVICE AVAILABILITY IN TERMS OF %TIME NETWORK PACKET LOSS (AT 10 SEC INTERVALS) IS BELOW A CERTAIN THRESHOLD

According to Table III, Internet2 clearly has the best performance, in that 97.7% of the time (in 10 sec intervals) are free of packet losses. Internet2+ performance is also good, but is noticeably lower than Internet2.

It should be noted that many of the packet losses are not evenly distributed. Therefore on many paths, especially international or commercial Internet paths, high loss and/or delay can sometimes be prevalent during an entire call. This is partly reflected in Table III as lower availability values for these types of paths.

Finally, in Table III, when the loss threshold is increased to 20%, we find that all types of networks achieve essentially the same level of availability, around 99.75%. We will see in next section that this is partly due to existence of outages (long losses).

Network/path type	%time loss is below			
	0%	5%	10%	20%
All	82.50%	97.60%	99.37%	99.91%
Commercial	78.85%	96.84%	99.27%	99.91%
Internet2	97.98%	99.89%	99.99%	99.99%
Internet2+	86.96%	98.53%	99.49%	99.91%
Domestic (US)	83.74%	96.95%	99.42%	99.93%
International	81.92%	97.91%	99.34%	99.90%
Domestic commercial	73.68%	94.87%	99.02%	99.89%
International commercial	81.53%	97.87%	99.39%	99.92%

TABLE IV

SIMILAR TO TABLE III, BUT AFTER REMOVING LONG OUTAGES

To be discussed in Section III-C shortly, outages are long loss bursts, which we define in this context as eight or more consecutive packet losses. Because in Table III, packet loss is computed every 10 seconds, a 1 sec long outage would result in a 10% average loss assuming no other losses during that 10 sec period. It is desirable to be able to distinguish between a more or less random 10% loss and an outage induced 10% loss. Table IV shows the percentage of time the packet loss is below is a certain threshold, after excluding outages. Every outages is treated specially as 100% loss for its duration. Therefore the distinction can be easily made between isolated losses and outages. Then we exclude outages (in both denominator and numerator) from the calculation in Table IV. However, we can see that its values are still very close to that of Table III. This suggests that most packet losses corresponding to outages do not occur in isolation, otherwise we would have seen a big shift between Table IV and III. This suggestion will be supported in Section III-C with the observation that a small number of outages are very long and make up most of the outage time. The only big change in Table IV is for Internet2 paths, where after removing outages, 99.99% of the time (compared to 99.79%) have periods with loss $\leqslant$ 20%. This suggests most of Internet2's losses are outages, as opposed to isolated losses.

Apart from outages, we have found packet losses in our measurement to be easily detectable. Since FEC is very useful in loss recovery, it is important to know how quickly the receiver can detect the advent of a lossy period. In our data, usually the lossy period lasts fairly long and covers an entire VoIP call. If the lossy period is shorter than the entire call, the transitions between good and lossy periods are usually distinctive. Figure 1 shows such an example.

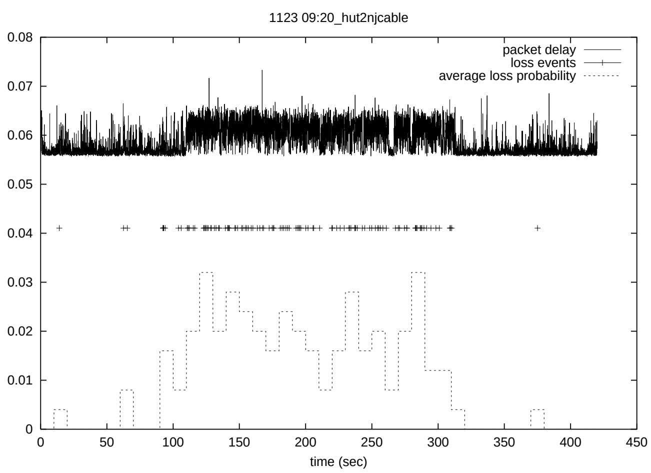

Fig. 1. Transition of packet loss conditions. For trace between node hut and njcable, Nov 23, 2002, 09:20

In Figure 1, the average loss probability is computed for every 10 seconds, as in Table III. Here losses become intense at time of 90 sec, and fades away at time of 320 sec. By observing the timing of loss events (the height is irrelevant), we can see a distinctive transition between path lossiness. In fact, the transition involves a visible change in delay as well. This means an adaptive application can adjust its FEC redundancy timely to maintain a good level of quality. It should be noted that the average loss probability can vary significantly within the same lossy period, for example from 1% to 3.5% in Figure 1. Therefore, an adaptive algorithm should be conservative by slightly over-estimating average loss. A detailed analysis of FEC adaptation efficiency is part of our future work.

C. Network Outages

Network outages are long, consecutive loss events, and are an important characteristics of the Internet. In this paper we will use a simple threshold of eight or more consecutive packet losses to determine whether a loss event belongs to an outage. At 40 ms packet interval, an outage will cause at least 320 ms of interruption, which is a noticeable degradation in speech in itself. Under this definition, of the 0.56% overall packet losses described in previous section, 23% of them are outages, making up a non-negligible portion. The percentage of packets belonging to outages is $23\% \times 0.56\% = 0.13\%$, and in Sec.III-B we observed that even at 20% loss tolerance, there are 0.25% of the time the network is worse than that. The outages clearly makes up a significant part of this 0.25%. Table V compares overall network loss with outages, and although different networks have quite different overall loss, the % of time they are in an outage is similar. This is an indication that the cause of outages is probably something beyond network congestion and network type, such as software bugs in router and link failure [5].

Most of the outages are short, but a small portion of them are extremely long and make up the majority time of all outages. Table V shows this phenomenon. For example, across all calls, the median outage duration is only 25 packets (1 sec), but the mean is 145 packets or 5.8 sec, and outages longer than 1000 packets (40 sec) occupy more than half of the total durations. The same property applies to nearly all types of networks. This means the Mean Time To Restore (MTTR) can sometimes be very high. Figure 2 illustrates the distribution of outages for all calls vs. those on Internet2, and for US domestic vs. international calls. When the complementary CDF is plotted in log scale, an exponential distribution will appear as a straight line. But in Figure 2 this is clearly not the case. The initial steep drop of the curves indicates that most outages are short. The sudden cut-off at around 400 sec is because the call duration in most of our measurements is 7 minutes (420 sec), with a small portion of calls to be 3 minutes long.

Another interesting aspect of outages is that they often occur in symmetry. That is, if $A \rightarrow B$ has an outage, $B \rightarrow A$ may also experience an outage, at almost the same time. Table V shows that 30% of outages across all calls are symmetric, and this is more prevalent in the commercial Internet. The information below illustrates this effect. For example, from node hut to suna there was an outage 37 packets long at time 194.40 sec plus 01:10 hours (EST), and a matching (symmetric) outage occurred in the reverse direction at time 194.64 sec.

```
Outages at Dec 5, 2002 01:10, Eastern Standard Time
01:10_hut2suna 37 from 194.40s; 31 from 204.96s;
01:10_suna2hut 103 from 194.64s; 19 from 205.12s;
01:10_suna2wustl 66 from 194.64s; 65 from 205.12s;
01:10_wustl2suna 37 from 194.44s; 31 from 205.04s;
```

Finally, the outages are often spatially correlated near the edge of the network. If $A \rightarrow B$ has an outage, then for some other node X either $A \rightarrow X$ or $X \rightarrow B$ is very likely to have

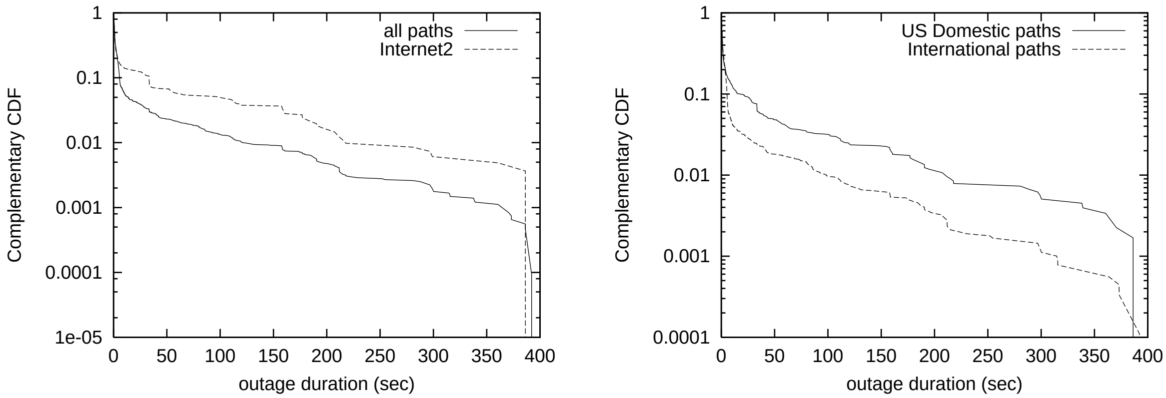

Fig. 2. Complementary distribution of outage durations

Network/path type	Overall loss	% time in outage	No. of outages	% symmetric	duration (packets) Mean	Median	total dur. (hh:mm) All	> 1000-pkt outage
All	0.56%	0.128%	10753	30%	145	25	17:20	10:58
Internet2	0.227%	0.183%	819	14.5%	360	25	3:17	2:33
I2+	0.422%	0.128%	2708	10%	259	26	7:47	5:37
Commercial	0.674%	0.129%	8045	37%	107	24	9:33	4:58
Domestic (US)	0.606%	0.123%	1777	18%	269	20	5:18	3:53
International	0.540%	0.131%	8976	33%	121	26	12:02	6:42
Domestic commercial	0.873%	0.08%	958	21%	190	15	2:02	1:20
International commercial	0.571%	0.154%	7087	39.5%	96	25	7:32	3:38

TABLE V
DISTRIBUTION OF OUTAGES

an outage at a similar time. And it is usually clear whether the outage is closer to node A or B. For example, in the above log information, the node suna is evidently closer to the problematic edge. A quantitative comparison would, however, require much more detailed analysis as well as intelligent filtering of data, which is part of our future work. Although the outages tend to occur at the edges, they usually stay only on one network, e.g., either Internet2 or commercial Internet. This is expected because these networks' infrastructures are completely independent.

D. Outage-induced Call Abortion Probability

During an outage, the user will hear a glitch and interruption of speech. If the outage lasts long enough, the user would eventually assume the call is dead and hang up the phone. The ITU E.855 standard [7] describes this behavior as call abandonment or call abortion. E.855 also includes a survey result revealing the following relationship between call abortion probability and the duration of an interruption (outage):

$$Pr[holding] = e^{-t/17.26} \tag{4}$$

where $Pr[holding]$ is the probability of a user still holding (not aborting) the call after the interruption has lasted t sec. Under Eq. 4, half of the users will abort the call after 11.96 sec of interruption. Using this model, we analyzed the expected (average) number of calls aborted due to these outages. If a call contains multiple outages, the holding probability is multiplied with its previous value. We also assume both sides may decide to hang up the call if he/she observes an outage. Across all calls, 2,566 of them have at least one outage, and among the 2566 calls the expectation (average) of the number of dropped calls is 946 under the E.855 model. This results in an overall call abortion probability of 946/62027 = 1.53%. Recall from Sec. III-A that the initial call failure probability is 0.47%, then the *net* service availability (meaning a call is successful and never aborted) becomes 100-(0.47+1.53) = 98%. In Sec. III-C the outages only make up 0.13% of all packets across all calls, but the call abortion apparently has some "amplifying" effect. Therefore, call abortion is an important user behavior that can result in visibly lower availability.

Network/path type	Call abortion probability
All	946/62027 = 1.53%
Internet2	99/8505 = 1.16%
Internet2+	314/27320 = 1.15%
Commercial	632/34699 = 1.82%
Domestic (US)	197/19858 = 0.99%
International	749/42165 = 1.78%
Domestic Commercial	98/11353 = 0.86%
International Commercial	535/23345 = 2.3%

TABLE VI
CALL ABORTION PROBABILITY WITH RESPECT TO NETWORK/PATH TYPE

Table VI shows the results for all types of networks. The international public (commercial) Internet calls have the highest call abortion probability, presumably because there are more outages on these paths.

IV. Related Work

Although there are a number of QoS studies on the Internet, we are not aware of any that studies VoIP service availability. Markopoulou *et al.* [10] discuss the performance of VoIP in the US backbone networks. Their focus is on quality as opposed to availability, and the networks measured are the backbone, inside US, as opposed to edge networks and internationally. In [9] Li also analyzes VoIP performance, but the focus is on quality, and the measurement points are located in research networks only.

Borella [4] analyzes loss patterns on the Internet measured among three test nodes. His results also indicate that a small portion of long loss bursts (defined as outages in our context) make up a significant portion of overall packet losses. However, the measurements in [4] do not have as much geographical, temporal (call-hours) and network diversity as ours. And his focus is not on service availability, but on other properties such as loss dependency and asymmetry. Bolot [2] and Paxson [11] also study packet losses in the Internet, with Bolot focusing on UDP loss dependency and Paxson on that of TCP.

The NLANR Active Measurement Project [6] also provides loss measurement among a large set of nodes, but the probing interval is relatively coarse grained at an average of every 60 seconds, which is too large to capture most of the outages that were detected in our measurements. Also, most of the NLANR test nodes are located on Internet2.

Finally, a recent feature article on ISP backbone reliability at NetworkWorld [7] is probably most relevant to our study. They measured packet loss behaviors including outages between test nodes from seven Tier-1 ISPs. The main difference of their study is their focus on backbone behavior, and the test nodes are located in the US. Also, the participating ISPs have a notion of maintenance window, some of which can occupy nearly 10% of the time. Although the measured reliability look very good (some exceeding five 9's), not counting such a large maintenance window for reliability calculation seems to "lower" the bar of availability at the definition level.

V. Conclusions

In summary, we have studied the service availability of VoIP achieved on the current Internet. Using our measurement data, we have analyzed several metrics of availability, including call success probability, % time loss is below a threshold, distribution of network outages and the call abortion probability induced by the outages. Overall, we observe that the call success probability at around 0.5% and call abortion probability at about 1.5%, resulting in a 98% net availability, which is still some steps away from what the PSTN offers today (three to four 9's), but already comparable to the availability of mobile telephone networks (around 97% to 99%). We plan to analyze our data in more detail and reveal further insights into the measurement results, such as node/edge-specific behaviors, characteristics of congestion-induced loss and delay apart from outages, effectiveness of FEC with adaptation in dealing with packet loss, and certain delay behaviors.

[6] http://moan.nlanr.net

[7] http://www.nwfusion.com/research/2002/1216isptest.html

References

[1] Gary Audin. Reality check on five-nines. *Business Communications Review*, pages 22–27, May 2002.

[2] J. C. Bolot. Characterizing end-to-end packet delay and loss in the Internet. *Journal of High Speed Networks*, 2(3):305–323, 1993.

[3] J. C. Bolot, S. Fosse-Parisis, and Donald F. Towsley. Adaptive fec-based error control for interactive audio in the Internet. In *Proceedings of the Conference on Computer Communications (IEEE Infocom)*, New York, March 1999.

[4] Michael S. Borella. Measurement and interpretation of Internet packet loss. *Journal of Communications and Networks*, 2(2):93–102, June 2000.

[5] Catherine Boutremans, Gianluca Iannaccone, and Christophe Diot. Impact of link failures on voip performance. Technical Report 200215, EPFL, Lausanne, Switzerland, April 2002.

[6] Qualex International. Trends in telephone service. Technical report, Federal Communications Commission, 445 12th St., S.W., Washington, DC 20554, May 2002.

[7] International Telecommunication Union. Connection integrity objective for international telephone service. Recommendation E.855, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, 1988.

[8] Wenyu Jiang and Henning Schulzrinne. Comparison and optimization of packet loss repair methods on voip perceived quality under bursty loss. In *Proc. International Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV)*, Miami Beach, Florida, May 2002.

[9] Fengyi Li. Measurements of voice over IP quality. Master's thesis, Department of Microelectronics and Information Technology (IMIT), Royal Institute of Technology (KTH), Stockholm, Sweden, June 2002.

[10] Athina Markopoulou, Fouad Tobagi, and Mansour Karam. Assessment of voip quality over Internet backbones. In *Proceedings of the Conference on Computer Communications (IEEE Infocom)*, New York, New York, June 2002.

[11] V. Paxson. *Measurements and Analysis of End-to-End Internet Dynamics*. PhD thesis, University of California at Berkeley, Berkeley, California, May 1997.

[12] J. Rosenberg and Henning Schulzrinne. An RTP payload format for generic forward error correction. RFC 2733, Internet Engineering Task Force, December 1999.

[13] J. Rosenberg, Henning Schulzrinne, G. Camarillo, A. R. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler. SIP: session initiation protocol. RFC 3261, Internet Engineering Task Force, June 2002.