Analysis of On-Off Patterns in VoIP and Their Effect on Voice Traffic Aggregation

Wenyu Jiang, Henning Schulzrinne
{wenyu,schulzrinne}@cs.columbia.edu
Department of Computer Science
Columbia University

Abstract—**We present an experimental analysis of on-off patterns in Voice over IP (VoIP), where we study the talk-spurt/gap distribution produced by two modern silence detectors: ITU G.729 Annex B Voice Activity Detector (VAD) and NeVoT Silence Detector (SD). The results indicate that spurt/gap distributions are fairly sensitive to both the sound volume and the type of silence detectors, but all of them showed that the traditional assumption of exponential distribution does not always fit well with the audio sessions we recorded. Both the spurt and gap distributions are more "heavy-tailed" than the exponential curve. In particular, the gap distribution deviates much more strongly from the exponential model, even when "hangover" is applied.**

To estimate how such deviation affects VoIP applications, we investigate the performance of voice traffic multiplexing. In particular, we look at the probability of having a out-of-profile packet (p_o) when a token bucket filter is placed at the multiplexing end. We run a series of simulations under three increasingly accurate settings: the exponential model, the real CDF, and the raw silence detector outputs. In general the token bucket results are fairly robust with regards to the details of the distribution. This is particularly true when the multiplexing factor N (number of voice sources) is large and the token buffer size B is not too big. When N is small and/or B is big, however, the estimated p_o under the real CDF is about 30% to 200% larger than under the exponential model. Finally, the relative difference between the raw silence detector outputs and the real CDF is generally much smaller than between the real CDF and the exponential model. Therefore, the data traffic in VoIP has a small temporal correlation and a secondary effect on the performance of multiplexers.

Keywords—**VoIP, IP telephony, traffic aggregation, QoS, on-off patterns.**

I. INTRODUCTION

Human speech consists of talk-spurts and silence gaps, also known as on-off patterns. The existence of spurts and gaps allows for silence suppression, where a voice segment is transmitted only if it is detected as active (a talk-spurt). The main benefits of silence suppression are:

- allows higher bandwidth utilization through multiplexing.
- allows per-spurt playout delay adjustment [17], [14].
- enable echo suppression based on silence detector output.

We are mainly interested in how much bandwidth utilization gain can be achieved by multiplexing, and what packet loss rate is introduced by the multiplexer for a particular utilization gain. Previous studies on the performance of voice traffic multiplexers [5], [15], [7], [11], [20] assume that the length of spurts and gaps follow an exponential distribution [2], [3], [4]. Since most of these speech measurements are based on either analog or simple digital silence detectors [2], [3], [4], [8], we suspected that the spurts/gaps produced by modern voice codecs and silence detectors will no longer fit well to the exponential model, which may in turn affect the packet loss rate at the same utilization gain.

We recorded several telephone conversations as digitized audio files. Next we applied to the audio files with G.729 Annex B Voice Activity Detector (VAD) [13] and the NeVoT [19] Silence Detector (SD). The resulting spurt/gap distributions to a large extent depend on the type of silence detectors and the volume level. In most cases, the spurt distributions is slightly more "heavy-tailed" than exponential, whereas the gap distribution deviates strongly from an exponential model.

We then run a set of simulations to study the effect of real spurt/gap distributions on multiplexer performance. A program simulates a token bucket with N on-off voice sources. Its token rate is expressed as a percentage R of the peak rate. It has a bucket depth of B (in counts of packet tokens). The performance parameter we examine is p_o, the probability that a packet is out of profile. The simulation results indicate that the exponential model in general gives a close estimate of p_o, particularly for a large N. In certain settings, however, the exponential model will under-estimate p_o by a large ratio.

The rest of the paper is organized as follows: section II describes the setup of telephony devices used to record telephone conversations. Section III compares the two silence detectors used in our experiment, G.729B and NeVoT SD. Section IV presents the spurt/gap CDF plots obtained from the silence detector outputs. Section V describes the token bucket simulation setup and its results.

II. EXPERIMENT SETUP

We used a gateway-based setup to record telephone conversations. As shown in Figure 1, it consists of a SIP-based [9] 3Com ethernet phone and a Mediatrix gateway, which is a 1-line PSTN-to-IP telephony gateway. The Mediatrix gateway performs a 2-wire to 4-wire conversion when it translates between PSTN signals and IP packets.

We record voice packets using "*tcpdump*". The dump file is filtered to retrieve the μ-law encoded RTP payloads, which are then stored as Sun μ-law ".au" files.

This work is supported by research grants from Hewlett Packard Labs

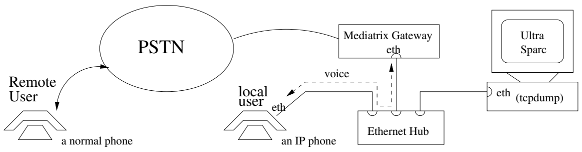


Fig. 1. Gateway-based telephone recording setup

III. SILENCE DETECTORS

A. Introduction to G.729B and NeVoT SD

We examine two silence detectors: G.729 Annex B VAD (Voice Activity Detector) [13], and the NeVoT Silence Detector (SD) [19]. Both of them use the energy in a voice frame as a first estimate in silence detection. NeVoT is based on the ISI VT audio tool [18]. The NeVoT SD uses a threshold that is dynamically updated but constrained with a min and max value. It uses a small hysterisis as well as a fixed but configurable "hangover" time. A hangover is a technique to avoid sudden *end-clipping* of speeches and to bridge short speech gaps such as those due to stop consonants. Within the hangover time, even a future silent frame is considered part of the latest talk-spurt. If any future frame within the hangover time is detected as active, the hangover time is "renewed". A similar technique is called *fill-in*, but it bridges a gap either in entirety or none, depending on whether the gap is shorter than the fill-in time. The fill-in time (typically 200 ms) introduces a significant look-ahead delay, making it unsuitable for telephony applications.

NeVoT SD has several configurable parameters:

Parameter	Meaning	Default
min thresh	frame energy below which any signal is considered silence.	-45 dB
max thresh	highest allowed silence threshold	-20 dB
pre	pre-spurt hangover time	1 packet
post	post-spurt hangover time	6 packets

As we will see in section IV, the min threshold and the total hangover are the more important parameters.

In contrast, G.729B's algorithm is more sophisticated, and its hangover time is not fixed. G.729B is also *fully automatic*. It does not require the user to set any threshold.

It is also noted that the G.729 Annex B spec uses a different volume measure than NeVoT SD. NeVoT SD uses a default min threshold of -45 dB, whereas the G.729B min threshold is -55 dB in the NeVoT volume scale and 15 dB in the G.729B scale. Therefore by default NeVoT SD is less sensitive than G.729B, that is, it tends to pick up less number of segments as talk-spurts. For the remainder of this paper we will use the NeVoT scale.

B. Comparisons with Traditional Silence Detectors

Traditional silence detectors such as those used by Brady [2] usually has fixed energy thresholds and fixed hangovers or fill-ins. Depending on the hangover or fill-in time (both denoted by T), the mean spurt and gap length can fall into two regions. If T is 0 or very small, mean spurt is around 200 to 400 ms, and the mean gap is around 500 to 700 ms. If T is around 200 ms, most short gaps are eliminated, and both the mean spurt and gap will be on the order of 1 to 2 sec.

Sriram and Whitt [20] quote[1] a mean spurt of 352 ms and a mean gap of 650 ms. Apparently this correspond to 0 or a small hangover. The ITU P.59 [12] recommendation specifies an artificial on-off model for generating human speech. It specifies a mean spurt of 227 ms and a mean gap of 596 ms without hangover, and a 1.004 sec and 1.587 sec respectively with hangover. Brady [4] gives an average of around 1.2 sec for spurts and 1.8 sec for gaps after applying hangover.

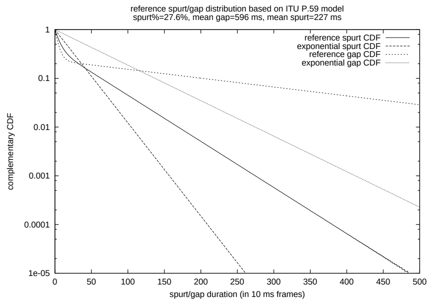
The G.729B VAD uses a dynamic hangover time. In fact, there are one frame long (10 ms) talk-spurts in the G.729B output. We will see in section IV that G.729B produces a distribution more like by a traditional silence detector without hangover. NeVoT SD behaves like a traditional silence detector, but it has a dynamically updated threshold.

Past speech measurements have in fact indicated that gap distributions without hangover do not always fit well with an exponential model [3], [8], [12]. In particular, the ITU P.59 [12] spurt/gap model without hangover is not exactly exponential, as seen in Figure 2 (a). But there are still several issues: First, previous studies on multiplexing performance have assumed an exponential model irrespective of the length of hangover. For example, Sriram and Whitt [20] used a mean spurt of 352 ms and mean gap of 650 ms. This is apparently without hangover, and the distribution is therefore not exponential. Second, G.729B is different from either a silence detector with no hangover or with a fixed hangover, and no study has been performed on how G.729B's dynamic hangover affects spurt/gap distribution. Third, although a long hangover (e.g., 200 ms) helps eliminate end-clipping of talk-spurts, it is unnecessary for modern voice codecs like G.729B because G.729B employs a sophisticated VAD algorithm and dynamic hangover along with Comfort Noise Generation. The long hangover is for a large part used to make Time Assigned Speech Interpolation (TASI) [6] work better (less jitter, reduced signaling overhead, etc.). This requirement is now obsolete in today's packet switched networks, because when individual voice flows are aggregated and sent to the router, it does not matter how continuous the stream is. A short/dynamic hangover only helps conserve bandwidth and reduce congestion.

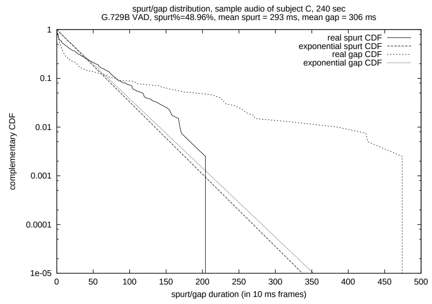
IV. CDF PLOTS OF SPURTS AND GAPS

Figure 2 (b) shows the complementary spurt/gap Cumulative Distribution Function (CDF) for one user in a recorded telephone conversation. In a *complementary* CDF the plot for an exponential random variable is a straight line when the y axis is in logscale. Therefore the two straight lines in Figure 2 (b) represent the *equivalent* CDF of spurts and gaps if they were exponentially distributed. Here the *equivalence* is

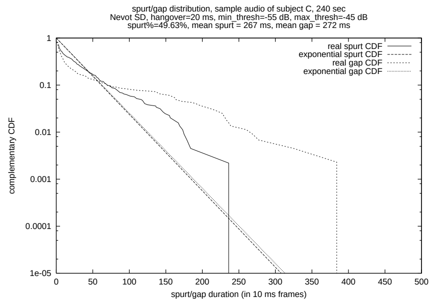
[1] They referenced it as a private work by May and Zebo, Bell Labs 1981.



(a) P.59 model, without hangover

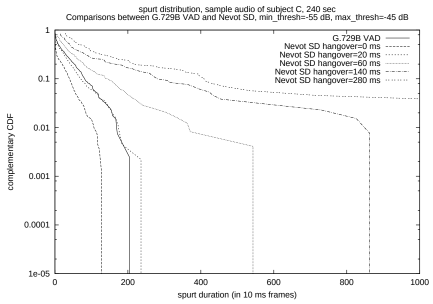


(b) CDF for one subject using G.729B

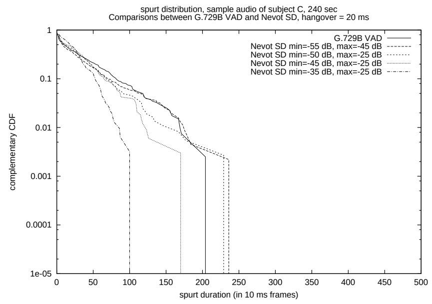


(c) The same subject using NeVoT SD

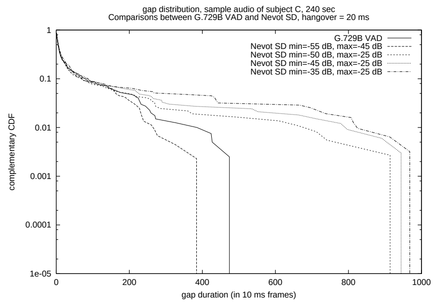
Fig. 2. Example spurt/gap distributions



(a) Spurt CDF using different hangover

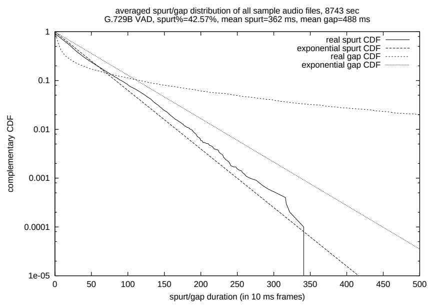


(b) Spurt CDF using different thresholds

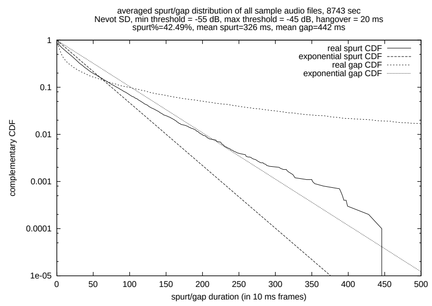


(c) Gap distribution using different thresholds

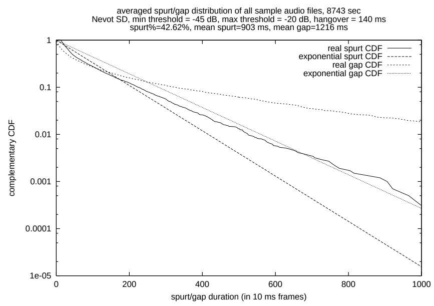
Fig. 3. NeVoT SD spurt and gap CDF using different parameters



(a) Averaged CDF by G.729B VAD



(b) Averaged CDF by NeVoT SD, over the same set of converstaions



(c) Averaged CDF by NeVoT SD with a high threshold and a large hangover

Fig. 4. Spurt/gap distribution after averaging over many converstaions,

defined as having the same mean value.

We recorded six conversations with an average duration of about 720 sec (the total time (8743 sec) printed at the top of CDF plot in Figure 4 divided by twelve). Five of the conversations were in Chinese, the other in English. We did not notice a visible impact of the language on spurt/gap distributions.

Figure 2 (c) is the CDF plot when the same audio file in Figure 2 (b) is run through the NeVoT SD. For this plot we choose the min threshold as -55 dB and a 20 ms hangover time, because it yields a similar peformance to the G.729B VAD. Figure 3 (a) is the CDF plot of the same audio file when varying the NeVoT SD hangover time. NeVoT by default uses a 20 ms frame, and a hangover of about 7 frames, therefore, it is equivalent to a 140 ms hangover time. We can see that a 140 ms hangover time can significantly change the CDF. NeVoT distinguishes between pre (default 1) and post-spurt (default 6) hangover. However, as far as the distribution is concerned, only the total number of hangover packets matters.

Figure 3 (b) is the CDF plot of the same audio file when varying the NeVoT SD min and max silence detecting threshold. The min threshold seems to be the most important factor.

The G.729B VAD is fully automatic, whereas the NeVoT SD has several configurable parameters. The most important ones are the min threshold and hangover time. Since the setting of these thresholds can have a significant effect on silence detection, we choose to use parameters that lead to similar performance to that of G.729 B. Therefore, we use a min threshold of -55 dB and a hangover of 20 ms.

Gap distributions are less sensitive to hangover time, but still sensitive to min threshold, as seen in Figure 3 (c).

Figure 4 shows the spurt/gap distribution produced by G.729B VAD when averaged over many conversations. Before averaging, the recordings are listened by the author and the sound volume is increased or decreased appropriately to minimize the effects of volume on silence detectors. We also tried to adjust the volume automatically, for example, by "normalizing" the average spurt energy to a reference dB value, but the resulting volume is still sometimes too loud or too weak. This is probably due to the difference in energy (dB) and loudness (subjective parameter).

We can see that the CDF plots are quite similar to that of Figure 2 (b). The spurt CDF curve is slightly above its exponential counterpart, which means it is slightly more "heavy-tailed". The gap distribution is significantly different from its equivalent exponential model. Therefore, we can conclude that the exponential model is apparently not a good fit for the gap distribution, and depending on the requirement, the exponential model may be considered an inadequate fit for the spurt distribution as well.

Figure 4 (b) is the equivalent plot of Figure 4 (a) for NeVoT SD. Its CDF is similar to that of G.729B VAD, although there is some difference in the mean spurt and gap length.

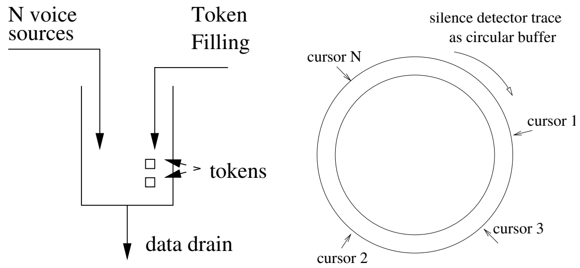
Figure 4 (c) is a similar plot when NeVoT SD uses its default setup (-45 dB min, -20 dB max threshold, 140 ms hangover). We can see that its mean spurt and gap are much longer, on the order of 1 sec.

V. Token Bucket Simulations and Results

A. Simulation Setup

Anick *et al* [1] gives an analytical procedure to derive the dynamics of a fluid producer/consumer system. Both the producers and consumers are on-off sources and sinks, respectively. Each of the M producers dumps fluid into a bucket at a fixed rate when it is in the on state, and sends nothing while in the off state. Each of the N consumers drains the bucket at a fixed rate when in the on state, and does nothing while in the off state. Both producers and consumers follow an exponential distribution in their on-off patterns, although with possibly different averages.

In VoIP traffic aggregation, a token bucket is usually used to perform multiplexing and shaping. Figure 5 is an example of a token bucket in action. The tokens are filled at a constant rate, and each packet consumes a token before it is transmitted. If there is no token available when a packet arrives, the packet is considered out-of-profile. It is up to the ISP to decide what to do with an out-of-profile packet. It can be either treated as best-effort, or discarded. Since the main performance indicator we examine is the out-of-profile probability p_o, the token bucket becomes equivalent to a leaky bucket with the same buffer size. The only difference is the queueing delay associated with a leaky bucket.



(a) A token bucket filter in action
(b) Illustration of trace-based simulation

Fig. 5. Token bucket VoIP multiplexer simulation setup

Bruno *et al* [5] used the results from Anick *et al* [1] to analyze VoIP aggregation with a token bucket. The voice sources correspond to the on-off consumers, and the token filling process correspond to the producers except that it is on all the time. Bruno's analysis also assumes the voice sources have exponential on-off patterns. They use a mean spurt of 350 ms and gap of 650 ms, about the same as in Sriram and Whitt [20]. Therefore, it also corresponds to spurt/gap without hangover, and hence not well fit to the exponential model.

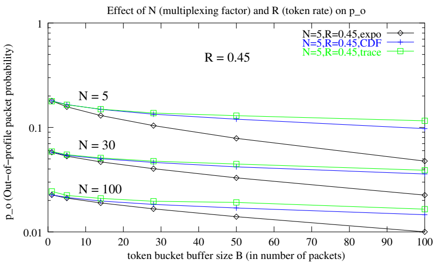
We have run a series of simulations that models the behavior of a token bucket multiplexer. The first set of simulations is based on exponential distribution[2]. The second set is based on the real spurt/gap CDF, obtained from our recordings of various telephone conversations. The last set is based on raw silence detector traces, that is, the raw output of either G.729B or NeVoT SD. This is to examine whether there is any temporal correlation effect that may influence the performance of multiplexers. The way we carry out a trace-based simulation is by creating a cursor (an array index) for each voice source. Each cursor is initialized to a random location in the silence detector trace, and traverses (and cycles upon the end) the trace sequentially from there on.

B. Results Based on the Exponential, CDF and Trace model

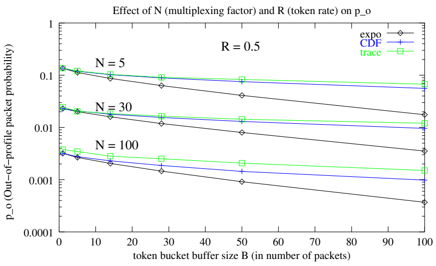
The parameters we used here are similar to the ones in Bruno *et al* [5]. Figure 6 shows the probability of an out-of-profile packet (p_o) for different multiplexing factors (N), token rates (R), and token buffer sizes (B). This set of simulation is based on CDFs and traces by the G.729B VAD. The unit of B is the number of packets. R is expressed as the ratio of the absolute token rate to the peak data rate. If on average a person talks 40% of the time, then R should be at least 0.4 to sustain the average data rate. In reality, R should be somewhat larger to absorb the burstiness of voice traffic when many sources are on and transmitting. The sample audio we used has an average spurt% (i.e., percentage of time in the on state) of 43% under the G.729B VAD.

From the plots we see that the exponential model generally under-estimates p_o by a small fraction. This under-estimate

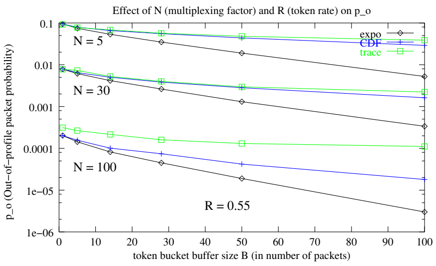
[2] Strictly speaking, it is *geometric* due to discrete packetization



(a) $R = 0.45$

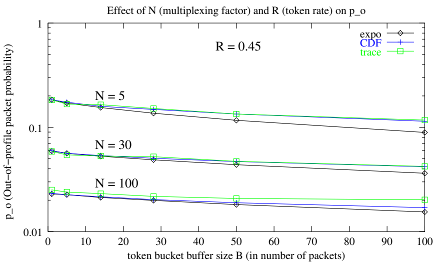


(b) $R = 0.5$

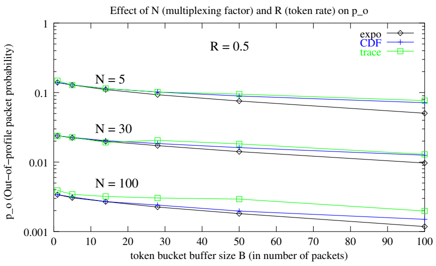


(c) $R = 0.55$

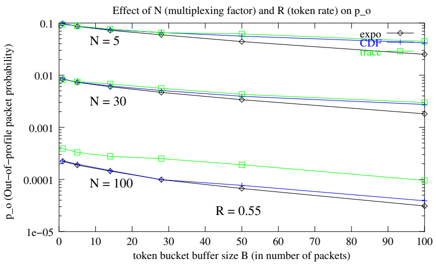
Fig. 6. Effect of spurt/gap distribution on multiplexing performance, G.729B



(a) $R = 0.45$



(b) $R = 0.5$



(c) $R = 0.55$

Fig. 7. Multiplexing performance for NeVoT SD with default parameters

is insignificant if the token rate R is relatively small (under-provisioned) and/or if N is large. When R is small, many packets will be out-of-profile, therefore the burstiness of CDF and raw trace data is less amplified in terms of p_o, because the base value of p_o will be fairly large. When N is large, p_o is likely to be small for the same R and B compared to a small N, therefore the absolute difference becomes negligible.

However, as seen in Figure 6, in certain cases, the relative difference of p_o between the exponential model and the CDF model can be quite large. Because Figure 6 has the ordinate in logscale, the distance between the curves at a given point B represents the multiplicative difference (ratio) or the relative difference. From Figure 6 we can see that the relative difference becomes very big for large B and/or small N.

Finally, the results from the CDF model also differs slightly from the trace model. This represents a small degree of temporal correlation in the spurt/gap traces compared to the memoryless CDF. This correlation consistently yields a higher p_o, which represents a burstier pattern than the CDF model.

Table I shows the numerical simulation results of Figure 6. Not all data points are listed. However, all data points with B = 100 are listed in Table I, to illustrate the strong deviation of CDF and trace based simulation from the exponential model. As seen in Table I, when token rate R is small, p_o will be quite large, therefore the slight difference in p_o between different models does not play an important role. E.g., when N=5, R=0.45, B=14, p_o is 0.13 under exponential model, and around 0.15 under the CDF or trace model. This difference is minimal given p_o is already quite high. This example may be somewhat unrealistic because people probably won't use VoIP if the loss rate is that high (assume out-of-profile packets are discarded). Another example is when N=5, R=0.55, B=100, p_o is 0.005 under exponential model, and around 0.03 under the CDF or 0.04 under the trace model. If the receiver application has no loss concealment [10], [16], a 0.5% loss could still be considered good quality, but a 3% to 4% loss is probably considered less as good as a 0.5% loss. The last example we consider is when N=100, R=0.55, B=100, p_o is 3×10^{-6} under exponential model, which is nearly perfect, but the CDF model gives 1.8×10^{-5}, 6 times higher. The trace model gives 1.11×10^{-4}, 37 times higher. It may be an extreme example, but it does show how big the relative difference can become. This data point also seems to be an anomaly point, because the trace-based results deviates strongly from even the CDF results. Such anomaly is also observed in Figure 7, when NeVoT SD with the default setting is used.

Figure 7 shows a similar set of performance plots. It uses the CDF and raw traces produced by NeVoT SD on the same set of audio files when it uses defaults parameters. That is, a min threshold of -45 dB, max of -20 dB, a hangover time of 140 ms. The plots looks similar to Figure 6, but the relative difference of p_o between the exponential model and the CDF is much smaller. This is probably because a large hangover makes the spurt/gap distribution closer to an exponential distribution. There also appears to be an anomaly point at (N=100,R=0.55,B=100). The trace-based p_o is consistently around 10^{-4} for both G.729B and NeVoT SD (large hangover) simulations at this data point. This is also true for simulations based on NeVoT SD traces with a small hangover. We do not know the cause of such anomaly, but it seems to indicate the sample audio trace can exhibit a strong temporal correlation in certain situations.

N	R .	B	p_o expo	p_o CDF	p_o trace
5	0.45.	14	0.130	0.149	0.150
5	0.45.	50	0.079	0.120	0.130
5	0.45.	100	0.048	0.097	0.116
5	0.5 .	14	0.087	0.102	0.104
5	0.5 .	50	0.041	0.075	0.083
5	0.5 .	100	0.018	0.056	0.067
5	0.55.	50	0.019	0.044	0.048
5	0.55.	100	0.005	0.029	0.039
30	0.45.	14	0.049	0.050	0.051
30	0.45.	100	0.022	0.035	0.039
30	0.5 .	28	0.012	0.015	0.016
30	0.5 .	100	0.004	0.010	0.012
30	0.55.	50	0.0013	0.0028	0.0030
30	0.55.	100	0.00034	0.0016	0.0022
100	0.45.	5	0.021	0.021	0.022
100	0.45.	100	0.010	0.015	0.017
100	0.5 .	50	0.00091	0.0014	0.0021
100	0.5 .	100	0.00037	0.00098	0.0015
100	0.55.	14	0.000082	0.00010	0.00022
100	0.55.	100	0.000003	0.000018	0.000111

TABLE I

SELECTED DATA RESULTS FOR SIMULATION

From a practical point of view, p_o will be small for a large B, therefore, even if the exponential model estimate is off by a large ratio, the aboslute difference is still small. For example, a user may or may not be able to tell between a 99.5% good circuit from a 99.0% good circuit. However, this difference may become important when stringent and precise traffic engineering is required, for example, when a company signs a contract with an ISP using a strictly specified Service Level Agreements (SLA). For an SLA, 0.5% loss and 1.0% could mean a significant difference.

VI. CONCLUSIONS

We present the analysis of on-off patterns (talk-spurts and gaps) for Voice over IP. We apply the G.729B Voice Activity Detector (VAD) and NeVoT Silence Detector (SD) to some recorded telephone conversations. The results indicate that spurt/gap distributions are not exactly exponential, particularly for gaps. The NeVoT SD can be tuned to behave similar to G.729B VAD with a comparable threshold and short hangover. We then conduct token bucket simulations based on the exponential model, the obtained spurt/gap CDF, and the raw trace of silence detector output. The performance indicator we examine is the out-of-profile probability (p_o). The simulation results indicate that the exponential model generally gives a close estimate of p_o, especially for large multiplexing factors. But the relative difference between these models can become quite large (about 30% to 200%) in certain settings, especially when the token buffer size is large. We have also observed an anomaly data point where the trace-based simulation result can deviate heavily even from the CDF-based result, which we suspect is due to some internal temporal correlation effect in the trace. In summary, the exponential model can be used for a first-hand performance estimate, but a more precise model (such as a CDF) is needed in certain settings and where high precision is required, for example when a strict Service Level Agreement (SLA) is to be determined.

REFERENCES

[1] D. Anick, Debasis Mitra, and M. M. Sondhi. Stochastic theory of a data-handling system with multiple sources. *Bell System Technical Journal*, 61(8):1871–1894, October 1982.

[2] Paul T. Brady. A technique for investigating on-off patterns of speech. *Bell System Technical Journal*, 44(1):1 – 22, January 1965.

[3] Paul T. Brady. A statistical analysis of on-off patterns in 16 conversations. *Bell System Technical Journal*, 47(1):73–91, January 1968.

[4] Paul T. Brady. A model for generating on-off speech patterns in two-way conversation. *Bell System Technical Journal*, 48(9):2445–2472, September 1969.

[5] R. Bruno, R. G. Garroppo, and S. Giordano. Estimation of token bucket parameters of voip traffic. In *IEEE ATM Workshop*, 2000.

[6] S. J. Campanella. Digital speech interpolation. *COMSAT Technical Review*, 6(1):127–158, Spring 1976.

[7] John N. Daigle and Joseph D. Langford. Models for analysis of packet voice communications systems. *IEEE Journal on Selected Areas in Communications*, SAC-4(6):847–855, September 1986.

[8] John G. Gruber. A comparison of measured and calculated speech temporal parameters relevant to speech activity detection. *IEEE Transactions on Communications*, COM-30(4):728–738, April 1982.

[9] M. Handley, H. Schulzrinne, E. Schooler, and J. Rosenberg. SIP: session initiation protocol. Request for Comments 2543, Internet Engineering Task Force, March 1999.

[10] Vicky Hardman, Angela Sasse, Mark Handley, and Anna Watson. Reliable audio for use over the internet. In *Proc. of INET'95*, Honolulu, Hawaii, June 1995.

[11] Harry Heffes and David M. Lucantoni. A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance. *IEEE Journal on Selected Areas in Communications*, SAC-4(6):856–867, September 1986.

[12] International Telecommunication Union. Telephone transmission quality objective measuring apparatus: Artificial conversational speech. Recommendation P.59, Telecommunication Standardization Sector of ITU, Geneva, March 1993.

[13] International Telecommunication Union. Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction annex b: A silence compression scheme for g.729 optimized for terminals conforming to recommendation v.70. Recommendation G.729B, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, November 1996.

[14] Sue B. Moon, Jim Kurose, and Don Towsley. Packet audio playout delay adjustment algorithms: performance bounds and algorithms. Research report, Department of Computer Science, University of Massachusetts at Amherst, Amherst, Massachusetts, August 1995.

[15] H. Naser, A. Leon-Garcia, and O. Aboul-Magd. Voice over differentiated services. Internet Draft, Internet Engineering Task Force, December 1998. Work in progress.

[16] Colin Perkins, Orion Hodson, and Vicky Hardman. A survey of packet loss recovery techniques for streaming audio. *IEEE Network*, 12(5):40–48, September 1998.

[17] Ramachandran Ramjee, Jim Kurose, Don Towsley, and Henning Schulzrinne. Adaptive playout mechanisms for packetized audio applications in wide-area networks. In *Proceedings of the Conference on Computer Communications (IEEE Infocom)*, pages 680–688, Toronto, Canada, June 1994. IEEE Computer Society Press, Los Alamitos, California.

[18] Eve M. Schooler and Stephen L. Casner. A packet-switched multimedia conferencing system. *SIGOIS (ACM Special Interest Group on Office Information Systems) Bulletin*, 10(1):12–22, January 1989.

[19] Henning Schulzrinne. Voice communication across the Internet: A network voice terminal. Technical Report TR 92-50, Dept. of Computer Science, University of Massachusetts, Amherst, Massachusetts, July 1992.

[20] Kotikalapudi Sriram and Ward Whitt. Characterizing superposition arrival processes in packet multiplexers for voice and data. *IEEE Journal on Selected Areas in Communications*, SAC-4(6):833–846, September 1986.