Analysis and Modeling of VoIP Conversation Traffic in the Real Network

Padungkrit Pragtong, Tapio J. Erke, and, Kazi M. Ahmed
Telecommunications Program, School of Advanced Technologies
Asian Institute of Technology (AIT)
Pathumthani, Thailand
st018481@ait.ac.th; erke@ait.ac.th; kahmed@ait.ac.th

***Abstract*— This paper presents the characteristic and traffic modeling of VoIP conversation. The traffic data are measured from the operating IP network of Telephone Organization of Thailand (TOT) Corporation. The observed distributions of talkspurt and silent durations considerably differ from the standard ON-OFF model (ITU-T P.59). It is proposed that a new state "long burst" representing the background noise at the talker's place, is added into the model based on continuous-time Markov process. The other two states, "normal speech" and "long silence", represent the normal behavior of the VoIP user. If background noise occurs during the speech, the model is classified as "noisy speech" otherwise as "noiseless speech". The simulation of traffic aggregation shows that the background noise significantly increases the data rate mean but reduces the variance. The measurement results support that both directions of the conversation can be modeled by the same model.**

***Keywords*— IP telephony, voice over IP (VoIP), voice traffic modeling.**

I. INTRODUCTION

Voice over IP (VoIP) or Internet Telephony transmits the human voice by using the Internet as a bearer. The voice signal is packetized into the IP packets and is reassembled at the destination. To guarantee the quality of VoIP service, the properties of VoIP traffic should be known.

The current VoIP traffic models are based on the conversational speech, in which the VoIP packets are generated during the on period (known as talkspurt duration, T_T) but not generated during the off period (known as silent duration, T_S). The exponential distributions of T_T and T_S are suited for English, Italian and Japanese languages [1]. However, in [2-4], it is shown that talkspurt and silent durations cannot be modeled with the exponential distribution in case of VoIP traffic.

This paper focuses on finding an appropriate model of talkspurt and silent duration of VoIP service. The traffic data analyzed in this paper are measured from the IP network of TOT Corporation. The TOT IP network and the measuring process are described in Section 2. The characteristics of VoIP conversation traffic are shown in Section 3. The VoIP traffic classification and the proposed models are illustrated in Section 4 and 5, respectively. Section 6 shows the verification of the proposed models and the effect of background noise on traffic aggregation.

This work was supported by the TOT Public Company Limited, Thailand.

II. VOIP TRAFFIC MEASUREMENT

The VoIP traffic data are measured from the IP network of TOT Corporation, which is the nationwide IP network of Thailand. The main services are VoIP and Internet dial-up. The VoIP service provides low cost call for long distance and mobile phone customers. The spoken language in the network is generally Thai. The voice codec is G.729A with payload size of 20 bytes, and the packetization delay of 20 ms.

The VoIP traffic is measured from Laksi (LKS) and Prakanong (PKN) switching centers in Bangkok, Thailand. The real-time analyzer and VoIP analysis software (Offline Mediapro) from RADCOM are used for measuring and analyzing the VoIP traffic. The real-time analyzer is connected to the switch, which is in-turn connected with the voice gateways. The measured traffic comes from one voice gateway in each switch, which is considered sufficient because each voice gateway is connected with four E1 (120 voice channels). The SPAN port feature in the Cisco switch [5] is used for forwarding the unicast traffic (in both directions) from the desired gateway port to the real time analyzer port. The connection for measuring the traffic data is shown in Fig.1

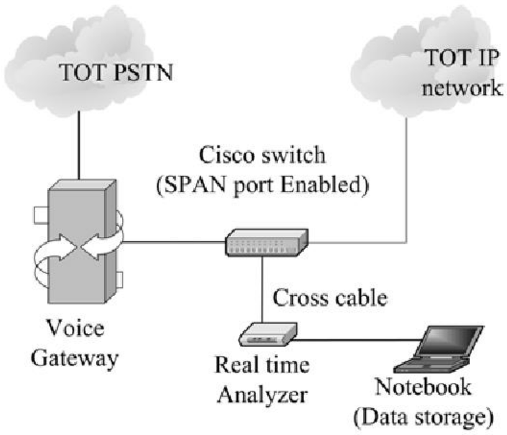


Figure 1. The connection for measuring the VoIP traffic

TABLE I. STATISTICS OF MEASUREMENT PROCESSES

Measurement	LKS	PKN
Dates (Year 2003)	3 - 29 April	28 May - 6 June
Number of calls	4,831 calls	2,639 calls
Total call duration	233 hours	89 hours

For finding the VoIP conversational speech model, all voice packets in the conversation are measured. Table I shows the statistics of the measurement processes. Due to the limited capacity of the captured data, many of the measured calls are not complete calls. For this reason, the VoIP calls that have the call duration less than 30 s will not be considered. After obtaining the VoIP traffic data, the characteristics of VoIP conversation traffic are described in the next section.

III. VOIP CONVERSATION TRAFFIC

In one VoIP conversation, there are two voice streams between the calling party (Subscriber A) and called party (Subscriber B). The analysis of Sub A and Sub B talking behavior in conversation is presented and then followed by the analysis of talkspurt and silent duration.

A. Behavior of Sub-A and Sub-B in VoIP Conversation

The behavior of the calling party (Sub A) and the called party (Sub B) are considered based on the average data rate. The calculation of average data rate is defined by (1)

$$Average\ data\ rate = \frac{Number\ of\ considered\ voice\ packets}{Call\ duration} \quad (1)$$

The pairs of average data rate of Sub A and Sub B from PKN and LKS switches are similar. The rates from PKN switch are shown in Fig. 2. The scatter diagram of observed rates of Sub A and Sub B pairs distribute roughly over range of 20-50 packet/s per partner, Fig.2. The distributions of Sub A and Sub B average rates in PKN and LKS switches shown in Fig. 3 differ only slightly.

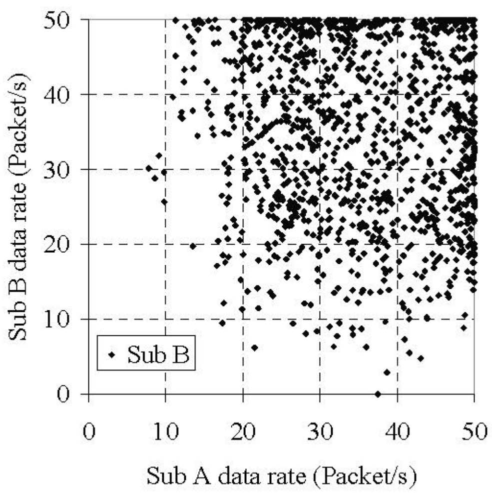


Figure 2. Sub A and Sub B data rate in 2,639 VoIP calls

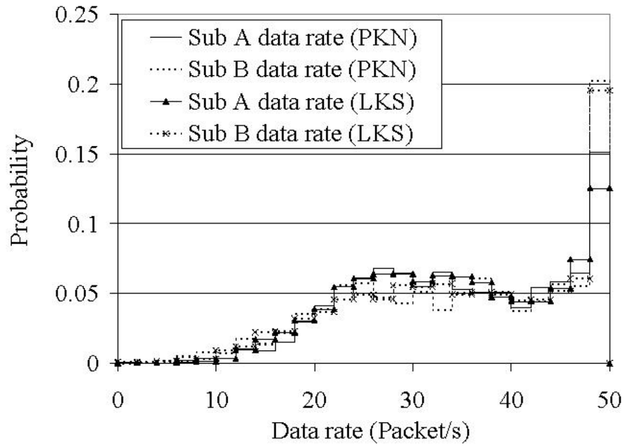


Figure 3. Distributions of average data rate of A- and B-subscribers

TABLE II. STATISTICS OF SUB A AND SUB B DATA RATE

Description	Sub A data rate (Packet/s)	Sub B data rate (Packet/s)
PKN average	35.11	35.14
PKN STD	10.37	11.64
LKS average	34.42	35.32
LKS STD	10.38	11.28

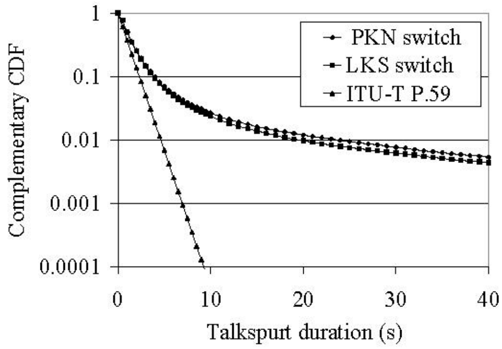
Table II shows that Sub A and Sub B data rates have the same characteristic. This implies that both directions of the conversation can be modeled by using the same model.

B. Talkspurt and silent duration

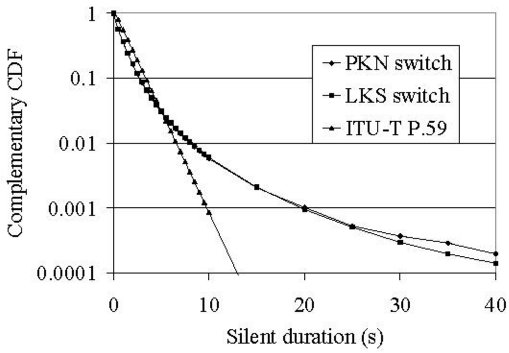
The observed distributions of talkspurt (T_T) and silent (T_S) durations are compared with the ITU-T P.59 distributions. The comparisons are based on the complementary cumulative distribution function (CCDF) as shown in Fig. 4. It is shown that the T_T and T_S from both switches are similar and have the long-tail characteristic. The observed T_T is longer than the longest possible actual talkspurt period in speech and cannot be modeled by the ITU-T P.59. For T_S, the result of standard model is close to the measured data but cannot track the tail of the distribution. The statistical comparisons of T_T and T_S between the measured data and standard model are shown in Table III. The results show that the average T_T in the measured data is longer than in the standard model. However, the average T_S in the measured data is shorter than in the standard model.

TABLE III. STATISTICAL COMPARISONS OF TALKSPURT AND SILENT DURATION BETWEEN THE MEASURED DATA AND THE PROPOSED MODEL

Statistics (s)	PKN	LKS	ITU-T P.59 [1]
Average Talkspurt	2.301	2.214	1.004
STD Talkspurt	7.510	8.242	1.004
Average Silent	1.184	1.182	1.587
STD Silent	1.881	1.871	1.387



(a) Talkspurt duration



(b) Silent duration

Figure 4. The comparison between the measured data and the ITU-T P.59

C. *Long-tail Characteristic of Silent and Talkspurt Duration*

The long-tail characteristic of T_S can occur because of the activities like holding the line for calling other people and listening to the called party. The long-tail characteristic of T_T happens because the call can be made from the place that contains a lot of background noise, e.g., from the street, in the cafeteria, etc. This noise causes the voice gateway to generate voice packets for longer durations because the Voice Activity Detector (VAD) in voice gateway cannot distinguish between the actual talkspurt and the background noise [6]. The duration of this talkspurt, voice plus background noise, is much longer than the normal speech. The causes for the long talkspurt duration are shown in Fig. 5.

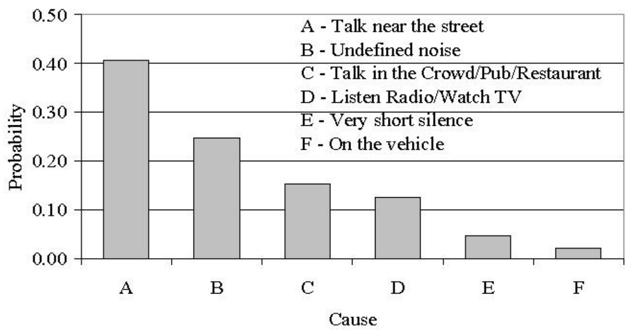


Figure 5. Causes of long talkspurt duration

In Fig.5, the most common cause of long talkspurt duration is the sound of traffic (talk near the street). This affects calls of mobile terminals and public payphones. Other causes are undefined noise, crowd, restaurant, radio or TV, vehicle, etc. Because there is no way to compress the background noise, it needs to be transmitted in the network to fulfill the conversation.

IV. Classification of VoIP Traffic

This section proposes the way to discriminate the background noise by using the concept of longest talkspurt (T_T) and silent duration (T_S) in conversational speech. The measured T_T and T_S are classified into states as described below.

A. *Normal speech*

The "normal speech" state contains two substates, which are the "talkspurt" state and "silence" states. Both of them represent the talkspurt ($T_{T-talkspurt}$) and silent ($T_{S-silence}$) durations of Thai speech, respectively. The averages of exponentially distributed Thai talkspurt and silent durations are 1.406 s and 1.110 s, respectively [7]. The longest $T_{T-talkspurt}$ is approximated as mean value plus four standard deviations, which for exponential distribution becomes equal to five times mean value. Consequently, the longest $T_{T-talkspurt}$ is approximately 7 s. Therefore, talkspurt durations, which are not longer than 7 s, are considered as talkspurt durations in normal speech ($T_{T-talkspurt}$).

For the longest $T_{S-silence}$, the parameter value in [7] cannot be used because it includes the silent duration that people listen to the call party. In [8], the longest silent duration in Thai monologue is found to be about 2 s. Therefore, the $T_{S-silence}$ must be less than or equal to 2 s. These silent durations represent the time that people need to take a breath before continuing their talk.

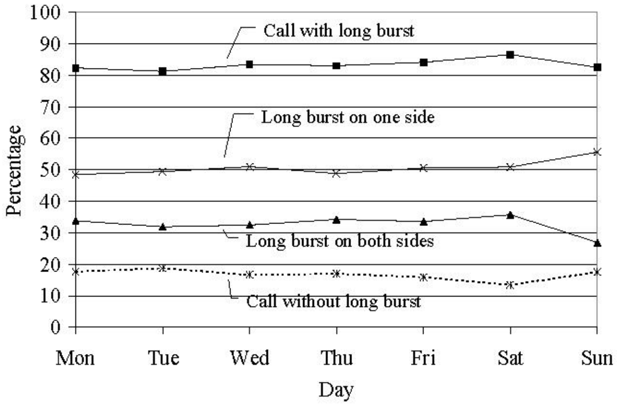
B. *Long silence*

The "long silence" state contains the silent durations (T_{S-Long}) that are longer than the $T_{S-silence}$ (i.e., longer than 2 s). The T_{S-Long} is caused by the activities like holding the line for calling other people.

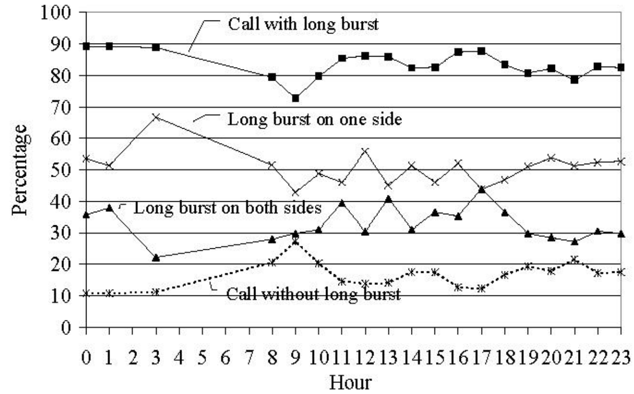
C. *Long burst*

The "long burst" state contains the talkspurt durations (T_{T-Long}) that are longer than $T_{T-talkspurt}$ (i.e., longer than 7 s). The T_{T-Long} can be found in the calls that are made from the place that contains a lot of background noise. This noise causes the voice gateway to generate voice packets for longer durations as mentioned earlier.

The occurrence of "long burst" state are considered. The results show that about 85 percent of calls in PKN and LKS switch have "long burst", and in other words, are affected by background noise. The occurrence of "long burst" state based



(a) Percentage of call with long burst based on day of the week



(b) Percentage of call with long burst based on time of the day

Figure 6. Percentage of call with long burst

on day of the week and time of the day in both switches are similar. The results from PKN switch are shown in the Fig. 6. It can be concluded that the occurrence of "long burst" in a call does not depend on the time of the day and day of the week.

V. THE PROPOSED MODEL

The proposed models are based on the continuous-time Markov process, where the distribution of state duration must be exponential. The proposed models are classified based on the occurrence of "long burst" state into the "noisy speech" and "noiseless speech" models. The "noisy speech" and "noiseless speech" models with transition rates are shown in Fig. 7. It is observed that the transitions between "talkspurt" state and "silence" state in both models occur frequently representing normal speech. However, the transitions from "long burst" state to "long silence" state in "noisy speech" model and vice versa occur rarely because it is difficult to have long silent duration in a noisy place.

VI. VERIFICATION

This section shows the verification of the proposed models based on the comparisons of state durations and data rate distribution from the measured data and the simulation of the proposed models and demonstrates the influence of background noise on aggregated traffic.

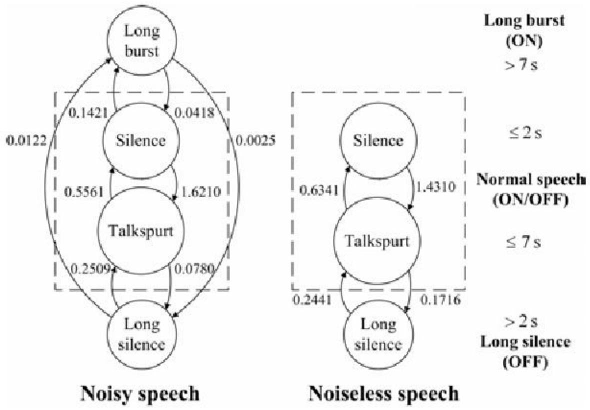
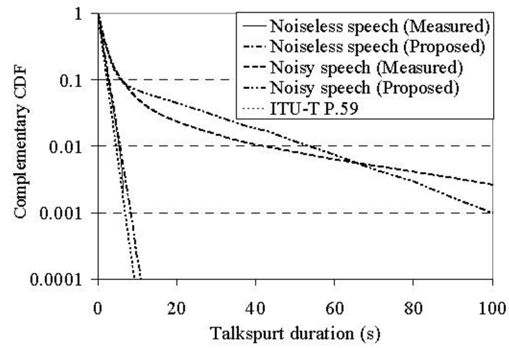
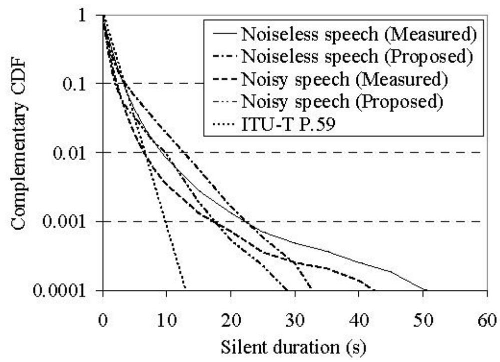


Figure 7. Transition rates in proposed conversational speech models



(a) Talkspurt duration



(b) Silent duration

Figure 8. Comparison of talkspurt and silent durations from the measured data and proposed models

A. Verification of talkspurt and silent duration

The comparison of T_T and T_S between the measured data and proposed models are shown in Fig. 8. It is observed that the simulation results of the "noiseless speech" and "noisy speech" models are reasonably close to the measured data. For talkspurt duration, the results of "noiseless speech" and ITU-T P.59 models are quite similar. For silent duration, the T_S in "noiseless speech" given results close to the ITU-T P.59 model, however, the tail of the proposed distributions can track the tail of the measured data better than in ITU-T model.

B. *Verification on VoIP Traffic Data Rate*

The comparisons of average data rate from the measured data rate and the proposed models are shown in Table IV. It is observed that the occurrence of background noise significantly increases the data rate of VoIP traffic. The data rates from the proposed models are close to the measured data and higher than the data rate of the standard model.

C. *Effect of Background Noise on VoIP Traffic Aggregation*

The proposed models are here used to demonstrate the influence of background noise on VoIP data rate generated by aggregated "noisy" and "noiseless" sources. The simulated network in Fig. 9 represents the corporate VoIP gateway connected with the 30 voice sources. The access network link is E1 (2 Mbit/s). This simulation is done by using the NS-2 simulator. The percentage of "noisy speech" sources is varied from 0 to 100 percent. The data rate on the access link indicates the aggregation of the VoIP traffic. The distribution of VoIP data rate of gateway traffic in 60 minutes observation period is shown in Fig. 10, together with aggregated data rate generated by ITU sources. Table V shows that increase in percentage of "noisy" sources significantly increases the data rate of traffic aggregation but decreases the standard deviation of the data rate.

TABLE IV. COMPARISON OF AVERAGE DATA RATE FROM THE MEASURED DATA AND THE PROPOSED MODELS

Description	Measured (Packet/s)	Proposed (Packet/s)
ITU-T P.59 [1]	19.38	
Noiseless speech	24.99	24.13
Noisy speech	40.91	41.61

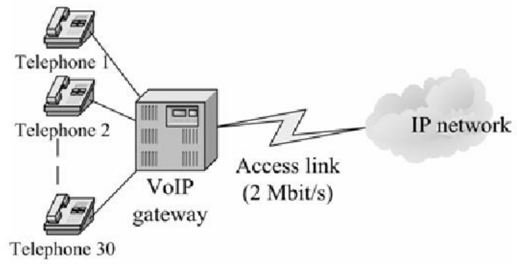


Figure 9. Simulation network

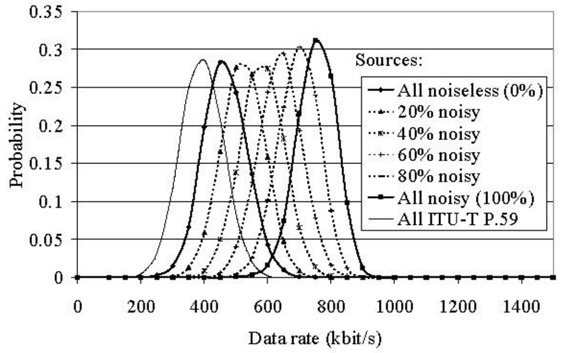


Figure 10. Distribution of aggregated data rates of 30 VoIP sources

TABLE V. STATISTICS OF AGGREGATED DATA RATES OF 30 VOIP SOURCES

Sources	Average (kbit/s)	Standard deviation
All noiseless (0%)	440.9	69.1
20 % noisy	496.8	67.0
40 % noisy	556.6	67.7
60 % noisy	614.2	65.9
80 % noisy	672.7	62.1
All noisy	728.5	59.6
All ITU-T P.59	367.98	64.30

CONCLUSION

The distributions of talkspurt (T_T) and silent (T_S) duration of conversational speech in the operating IP network cannot be modeled by the standard model because they have the long-tail characteristic. To understand the long-tail characteristic, the measured T_T and T_S are classified into three states as "normal speech", "long silence" and "long burst". The "normal speech" and "long silence" always occur in the conversation. The "long burst" represents the effect of background noise during the conversation. The most common cause of "long burst" state is the traffic noise. This affects calls of mobile terminals and public payphones. The "long burst" occurs in 85% of the calls.

The proposed Markov models are classified based on the occurrence of "long burst" as the "noisy speech" and "noiseless speech" models. The verification of the proposed models shows that the "noiseless speech" and standard models can be used for modeling the talkspurt and silent durations in the conversational speech without background noise. The proposed "noisy speech" model is used when the background noise is incorporated. The occurrence of background noise significantly increases the average but decreases the variance of the data rate of VoIP traffic and cannot be ignored.

REFERENCES

[1] Telephone transmission quality objective measuring apparatus: Artificial conversational speech, ITU Recommendation P.59, Mar. 1993.

[2] W. Jiang, H. Schulzrinne, "Analysis of On-Off Patterns in VoIP and Their Effect on Voice Traffic Aggregation", Computer Communications and Networks, Proceeding of Ninth International Conference, 2000., p.82-87

[3] P. Pragtong, K. Ahmed, and T. Erke, "The On/Off Model for Voice over IP Traffic", The Third International Symposium on Information and Communication Technology (ISCIT 2003), Vol.2, 670-674 pp., Songkhla, Thailand, September 2003.

[4] E. Casilari, H. Montes and F. Sandoval, "Modelling of Voice Traffic Over IP Networks", The Third International Symposium on Communications Systems Networks and Digital Signal Processing (CSNDSP 2002), Staffordshire, UK, July 2002.

[5] Cisco, "Configuring SPAN and RSPAN", November 2002. Available:http://www.cisco.com/en/US/products/hw/switches/ps700/products_configuration_guide_chapter09186a008007fa3a.html

[6] J. Davidson, and J. Peters, "Voice over IP Fundamentals", Cisco press, 2000.

[7] H. T. C. Hien, "Analysis of the Conventional Speech Patterns of Different Language Based Speakers over a Telephone Channel", Asian Institute of Technology, Master Thesis, Dec. 1995.

[8] S. Luksaneeyanawin, "Acoustic Characteristics and Functions of Speech Pause in Thai", Research Monograph, Linguistics Research Unit, Chulalongkorn University, Thailand, May 1988. (In Thai)