Characterizing Multiparty Voice Communication for Multiplayer Games

Gabor Papp and Chris GauthierDickey
Department of Computer Science
University of Denver
Denver, CO, USA
{pgabor | chrisg}@cs.du.edu

ABSTRACT

Over the last few years, the number of game players using voice communication to talk to each other while playing games has increased dramatically. In fact, many modern games and game consoles have added voice support instead of expecting third-party companies to provide this technology. Unlike traditional voice-over-IP technology, where most conversations are between two people, voice communication in games often has 5 or more people talking together as they play.

We present the first measurement study on the characteristics of multiparty voice communications. Over a 3 month period, we measured over 7,000 sessions on an active multi-party voice communication server to quantify the characteristics of communication generated by game players, including overall server traffic, group sizes, sessions characteristics, and speaking (and silence) durations.

Categories and Subject Descriptors

I.6.5 [**Simulation and Modeling**]: Model Development

General Terms

Measurement, Design

Keywords

Voice Communication, Computer Games, Talkspurts, Silence periods

1. INTRODUCTION

Our research characterizes multiparty voice communications over the Internet, and in particular when it is used with multiplayer games. To conduct our measurement study, we set up a TeamSpeak server which allows clients to join the server, set up channels, and communicate with other clients on the same channel [4]. In this client/server architecture, the clients encapsulate voice packets using the Speex codec [2], and send those packets to the server using unicast. The server then unicasts the packets to the other $n-1$ clients connected on the same channel, without multiplexing the voice packets. Table 1 lists the configuration parameters of our TeamSpeak server.

SIGMETRICS'08, June 2–6, 2008, Annapolis, Maryland, USA.

Table 1: Configuration parameters for TeamSpeak

Parameter	Value
server	TeamSpeak
port	8767
protocol	UDP
codec	12.3 and 16.3 kbps Speex

Once the TeamSpeak server was set up, we began logging all traffic on port 8767 to the server using *tcpdump*. After analyzing the data, we found that the incoming and outgoing voice packets all had one of the following four sizes: 183, 189, 233 or 239 bytes. Further, we identified the codec ID in the packets as well as the sender ID and a sequence number that ensured that we correctly filtered the voice packets. The difference between the different packet sizes is due to the bandwidth of the used codec and the fact that the outgoing packets contain slightly more information than the incoming ones.

As we began the analysis of our data, we discovered that some of the data points were extremely different from the rest. For example, some talking periods were several hours long without an intervening silence, even though we could detect 100ms silence periods. Some silence periods were days long, probably indicating that the server did not correctly detect the person logging out of the system. We linearized the data, identified the first and third quartiles, and removed any extreme outliers which were identified to be 3 times the inter-quartile range from either before the first quartile or after the 3rd quartile. Out of several hundred thousand packets, this process removed up to approximately 300 packets, depending on the data set.

In order to ensure that our data was not biased due to geographical location of clients connecting to the server, we estimated the location of our users based on their IP address. We found that more than 85% of our users were from the United States and more than 96% of them were from North America. This shows that our results are not biased by time zone or location within North America. Note, however, that most users will connect to a server that is within the same continent due to the effects of latency on voice communication.

2. MEASUREMENTS

2.1 Overall Server Traffic

The first set of measurements we present are the overall traffic seen by the server during an average day. This is the average number of packets that the server either received or sent out grouped by hours. In the busiest hour, we captured more than 20000 incoming and more than 100000 outgoing packets. Our server experienced the lowest traffic around 4am (less than 2000 packets in both directions).

Our results showed that the peak period is between 6pm-9pm and that the traffic rate during this time is almost constant. In other words, the number of sessions started is the same as the number of sessions finished during this period and thus resembles a balanced birth-death process. All the following measurements that we present in this paper are done during this period. The difference between the ratio of the amount of incoming and outgoing traffic is also the highest during the peak. The server input doubled and server output increased by an approximate factor of 5. This indicates that more users are online using multiparty voice communication during the evenings.

2.2 Group Sizes

We next examine group sizes to gain an insight into the size of a group that is typical in multiparty voice communication when used with games. As we noted previously, the ratio between the inbound and outbound traffic is an indicator of the average group size. We binned all data according to how many people it was duplicated to, allowing us to examine the data based on the size of the group. Thus, we could determine the effect of the groups with different sizes on both the incoming and outgoing traffic on the server. *The results on group sizes show that the most active groups are the ones that are formed by 5 people.* We also concluded that these groups generate the most outgoing traffic among all the groups. Groups of 2 people generated a similar number of incoming packets as 5 person groups (but generated less outgoing traffic due to replication). The largest group we observed was 24 people.

2.3 Sessions Characteristics

Over the period of measurement, we recorded 7,749 sessions, including the packets that were sent to and from the server and how long users were logged into TeamSpeak. *On average, we observed 86.1 logins per day from 721 individual users.* To understand this data further, we calculated the session times. Our calculations show that the shortest sessions were less than one second while the longest session was over 69 hours! However, for 20% of the sessions, users stayed less than 1/2 hour. In addition, 20% of the sessions, users stayed for more than 5 hours. Thus, 60% of the sessions fell somewhere between 1/2 hour and 5 hours. For the small fraction of sessions that were greater than 8 hours, we hypothesize that users simply did not log out of the TeamSpeak server when they were done. The characteristic of our curve is similar to work in [1] and [3].

In the next measurements, we matched IP addresses with sessions to determine how many unique IP addresses logged into the system. In essence, we wanted to determine how frequently a user logs into and uses the TeamSpeak server. Our results indicate that 50% of the users logged into the TeamSpeak server only once, while only 15% logged into it regularly. However, this result is most likely biased due to the fact that some users may be using DHCP to receive their IP addresses when they use the Internet. Thus, multiple IP addresses may refer to the same users and the total number of users we saw may be fewer.

2.4 Measured Voice Patterns

Voice patterns in multiparty voice communication consist of talkspurts (on periods) and silence (off periods). We measured these to characterize voice patterns. TeamSpeak uses 100ms long frames, therefore the shortest talkspurt in our case is 100ms. To be consistent, the smallest measurable silence period must also be 100ms. In order to measure the voice patterns, we captured 188,225 individual voice packets during the peak periods (6pm–9pm) during the three month measurement period.

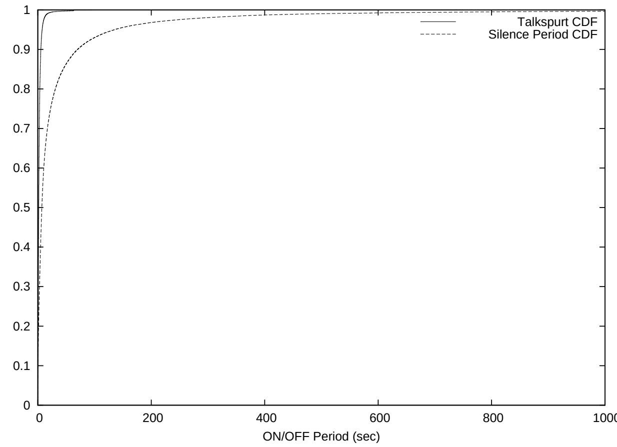

Figure 1: Both the CDF of the talkspurts and silence periods appear to follow an exponential distribution, where 90% of the talkspurts are less than 5.4 seconds. However, the expected value of the silence periods is much higher. The majority (90%) of the silence periods last less than 70.11 seconds.

In Figure 1, we plotted the CDFs of the talkspurts and silence periods Both CDFs appear to follow an exponential distribution. However, the expected value of the talkspurts is much lower than the expected value of the silence periods. *90% of the talkspurts are shorter than 5.4s, whereas 90% of the silence periods are shorter than 70.11s, which is more than an order of magnitude higher.* This implies that the users tend to listen more than they talk.

3. CONCLUSION

We have presented the first work that examines the characteristics of multiparty communication for games. While VoIP has been successful for point-to-point communication, and research has looked at the feasibility of VoIP over the Internet, our work is the first work to examine multiparty voice communications. Our results present an insight to the overall server traffic, the group sizes as well as the typical silence periods and talkspurts for multiparty voice communication, which can be used for future research, simulation, network engineering, and game development work.

4. REFERENCES

[1] T. Henderson and S. Bhatti. Modelling user behaviour in networked games. In *MULTIMEDIA '01: Proceedings of the Ninth ACM International Conference on Multimedia*, pages 212–220, New York, NY, USA, 2001. ACM Press.

[2] Speex Codec. `http://www.speex.org`.

[3] P. Svoboda, W. Karner, and M. Rupp. Traffic analysis and modeling for world of warcraft. *2007. ICC '07. IEEE International Conference on Communications*, pages 1612–1617, 24-28 June 2007.

[4] TeamSpeak. `http://www.teamspeak.com`.