Analysis of Call Length and Placement Times

Ian Fette
School of Computer Science
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213
Email: icf@cs.cmu.edu

***Abstract*— In this paper we examine the effect that the time of day has on phone calls made from a booth of three public telephones in Times Square, New York. Specifically, we look at the frequency with which these phones are used, and the length of calls placed on these phones, and how these factors vary with the time of day. We conduct our study by means of a natural observation facilitated by a webcam, and have observed that calls are both longer and more frequent between 8am and midnight than from midnight to 8am.**

I. Introduction

In this paper, we hypothesize that payphone calls are longer and more frequent from the hours of 8am to midnight, and shorter and less frequent from the hours of midnight to 8am. While this might seem a simple subject, the results do have many applications.

As an example, payphone companies may wish to know what times of day are the least active, so that they may perform maintenance at a time where the least number of users are inconvenienced. Or, long-distance companies may wish to know when the longest phone calls are placed, so that their rates may be set accordingly. These are but two possible applications of the data.

II. Methods

A. Subject Identification

This study was conducted via a naturalistic observation of three payphones in Times Square, in the heart of New York city. The payphones are all clustered together, and are on the corner of 46^{th} and Broadway streets. At least one of the phones is in use approximately 27% of the time, as measured over the entire day. We observed these phones from 4:43pm on October 7, 2005 to 4:44pm on October 11, 2005.

These observations were taken via a webcam streaming pictures over the Internet. Samples were taken every one minute. Over the four days studied, this amounted to approximately 5,242 images. (The camera experienced difficulty early Monday morning, and was off from 3:00 am to 8:20am.) The camera is available online[1] and is owned by EarthCam, Inc.

The bank of payphones observed are shown in Fig. 1. These phones are reasonably lit at all hours of the day, as well as at night by a number of street lights. We collected data for each of the three phones individually, and present such data in aggregate in this paper.

Fig. 1. Observed Phone Booths

B. Study Design

In this study, our main task was to identify whether a person was making a call on the payphones by looking at the collected image. Since we do not have access to the call logs from the phone company, this was done in a purely visual manner. While this task initially seemed simple, it became readily apparent that visually identifying a phone call in progress from images sampled at one minute intervals is not a trivial task. For each of the images, we attempted to apply the following criteria to classify an image as a phone call.

1) If a person is inside a phone booth, either wholly or in part, they are making a phone call
2) If a person is visibly holding a yellow phone headset, they are making a phone call
3) If a person is facing the phone booth and standing within arm's length, they are making a phone call
4) If a person is standing in front of a phone booth for multiple images and either (1), (2), or (3) is true for any image, then they are making a phone call for all images in which they stand within arm's distance of the phones

Unfortunately, the phones are not the main subject of the webcam. The phones are occasionally obstructed by large crowds of people, or people with large umbrellas. This poses a problem when trying to determine whether a phone call is being made. In cases where we could not make out a phone, we assume that the phone was in the same state (used or not used) as when we could last clearly make it out.

For the purposes of defining call length, we consider a single image in which a person is making a phone call (as defined above) to map to being on the phone for one minute. If a person is observed making a phone call for n images, then we consider the length of that phone conversation to be n images.

Defining call frequency is slightly more complex and very important given that there was a lack of data for a period of time. In this paper, we define call frequency as the number of calls begun within a fixed time divided by the number of observations made in the same fixed time. It is important to note that this is not the probability that a phone is in use at any time during that period. Rather, this most closely represents the probability that in any given minute during the specified period, a phone call will be placed (started).

C. Materials Used

To ensure uniform sampling every one minute, images from the webcam were retrieved and stored programatically. This program was a simple PHP script called by a crontab every minute. The system clock was set and maintained via NTP to ensure that there was not any clock skew which would affect the times associated with the images.

To analyze this large number of images, another PHP script was developed to show all of the images one at a time, on a web page. With each image was a form containing options for each phone corresponding to phone use or lack thereof. This facilitated the population of a MySQL database, which was used to store the analysis of all the images.

III. Results

A. Call Length

For analysis, we divide the day into six disjoint groups of four hours each. The length of phone calls as a function of the time of day at which the call was initiated is shown in Fig. 2. As can be clearly seen from the graph, the groups between 8am and midnight (points seen between 8 and 24 on the graph's x-axis) contain by far the most calls. If we envision the call day starting at 8am, calls proceed with diminishing average length until we hit approximately 6am, at which no calls are recorded. Indeed, the period between midnight and 8am appears characterized by infrequent and short calls. The periods after 8am appear characterized by more frequent and longer calls.

Upon initial examination of the data in four hour segments, it appeared that the average length of conversations was highest at the 8am-12pm block, and decreased strictly as time progressed towards midnight and again back to 8am. The averages shown in Table I would tend to support this, if not for the high standard deviation of call times. The averages are mostly inflated by outlying datapoints. If observed over a longer period of time, there might be more outlying datapoints suggesting a bimodal distribution, but for now it appears the only clear result is that these outlying calls happen much more frequently in the time from 8am-12am than from 12am to 8am.

We present in Table I not only the average length of phone calls within this group, but also the standard deviation. The

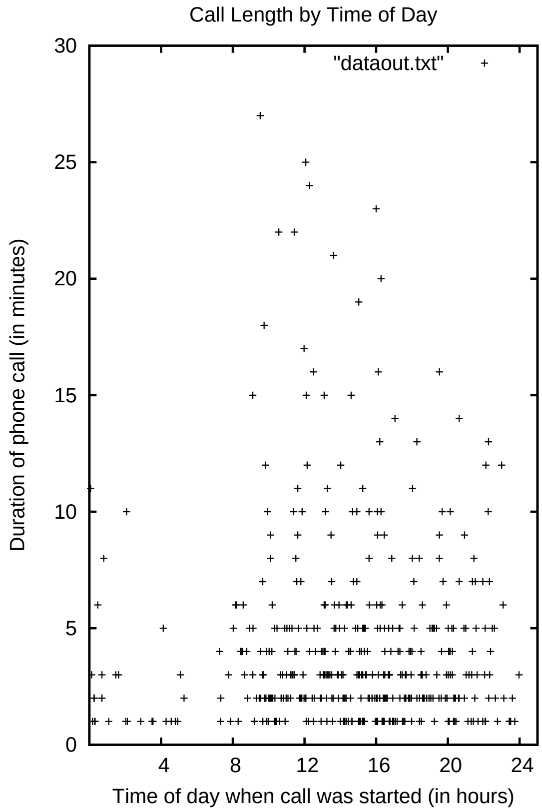

Fig. 2. Observed Phone Booths

TABLE I
PHONE CALL DURATION (IN MINUTES) FOR VARIOUS TIMES OF DAY

	8a-12p	12p-4p	4p-8p	8p-12a	12a-4a	4a-8a
Average	4.85	4.68	4.24	3.99	3.05	2.09
Std. Dev	4.84	4.58	3.96	3.23	3.06	1.38

high standard deviation is reasonable considering the large percentage of calls at five minutes or less, and the still considerable number of calls distributed between 6 and 27 minutes. The differences are not statistically significant between all pairs of time groups. The pairwise significance of each four hour group can be found in Table II. The pairwise significance was calculated using a One Sided Two Sample Kolmogorov Smirnov Goodness of Fit test[2], with the assumption that call length would decrease as the day progressed from 8am towards the nighttime hours.

A closer look at Table II reveals that the only statistically significant differences arise between groups from midnight to 8am and groups from 8am to midnight. If we reduce the number of groups to two, and set the group boundaries as 8am to midnight and midnight to 8am, we do indeed get a significance level of .011. While there is not a 95% chance of a significant difference between adjacent "daytime" groups, this may be due largely to the high number of outlying data points. Future work might look at the nature of these outlying

TABLE II

PAIRWISE P-VALUES FOR SIGNIFICANCE OF CALL LENGTH DIFFERENCES AT TIME OF DAY INTERVALS

	8a-12p	12p-4p	4p-8p	8p-12a	12a-4a
12p-4p	.569				
4p-8p	.446	.342			
8p-12a	.114	.308	.238		
12a-4a	**.036**	**.017**	**.054**	.093	
4a-8a	**.026**	.063	**.054**	.181	.575

TABLE III

FREQUENCY OF CALLS FOR VARIOUS TIME INTERVALS

	8a-12p	12p-4p	4p-8p	8p-12a	12a-4a	4a-8a
Frequency	.11	.13	.12	.07	.03	.02

data points. Unfortunately, we do not have enough analyzed data to make statements about the distribution solely of these outliers.

B. Call Frequency

As noted earlier, we define call frequency to be the number of calls begun within a fixed time divided by the number of observations made in the same fixed time. We present the call frequencies for the previously defined time intervals in Table III. Again we see the frequencies from 8am to midnight being relatively similar to each other, while markedly different from the hours of midnight to 8am.

IV. DISCUSSION

In this naturalistic observation study, we have shown that calls tend to be on average longer and more frequent between 8am and midnight than those calls between midnight and 8am. This seems to match with general intuition as to when people are most active. With that said, naturalistic observations have their limits, limits which greatly affect the generalizability of these data.

The first limit is the accuracy of the data. In our observations, we sampled the environment every one minute. This leaves us with the potential to miss calls shorter than one minute with increasing probability as the length of call approaches a small number of seconds. We also were not able to observe the phones at all times, and had to make inferences about times at which we could not see the phone. As far as accuracy of phone call time and duration recording, it would be greatly preferable to obtain logs from the phone company. That way we could be sure of the exact length of a call, and could differentiate between someone making a call and someone standing in a phone booth to avoid raindrops.

The second main limit is the number of confounders present in this study. This is an observation of a specific set of phone booths on a single corner in Times Square. Times Square is filled with life (and nightlife), and is hardly representative of payphones in the cornbelt, for example. At best, these results are generalizable to bustling city center areas.

Further confounders include activities taking place in this area at night that might not be so prevalent in other areas. For example, the images captured what looked like the arrest of prostitutes, as well as groups of people loitering, what might have been a drug deal, and what looks like one person about to inflict bodily harm upon another (this was the last image before the camera went out on October 10.)

Finally, some people seem to have been aware that they were being watched by a camera. it's not clear that this was an entirely covert observation. Indeed, many people appear to wave to the camera, and then make a phone call to somebody. While most of these calls were made on cell phones, there were a few individuals who waved and then used the payphones under observation.

With respect to the specific medium of webcams, there are a number of questions that remain unanswered. The quality of sampled data versus continuous data may have a large effect on the observed results. In this study, we assumed that we would have sufficient accuracy to at least make general statements. A study to see the difference in results from various sampling rates might prove insightful.

With all that said, we can conclude from the data that phone calls were made more frequently and with a longer average duration between 8am and midnight than from midnight to 8am for this specific set of phones, and possibly the other payphones in Times Square where there is much activity and multiple overt cameras. Any further generalizability is difficult, at best.

Future studies in the area would do well to look at a number of phones in more geographically diverse areas. We were limited primarily by time and resources in this study. Also, there is an inherit limitation to observations by webcams. Phone logs, while not perfect, would certainly be a better medium of observation.

REFERENCES

[1] I. EarthCam. (2005, Oct.) Times square. [Online]. Available: http://www.earthcam.com/usa/newyork/timessquare/index.php

[2] I. M. Chakravarti, R. G. Laha, and J. Roy, *Handbook of Methods of Applied Statistics*. John Wiley and Sons, 1967, vol. 1, pp. 392–394.