Understanding User Behavior at Scale in a Mobile Video Chat Application

Lei Tian[1], **Shaosong Li**[1], **Junho Ahn**[1], **David Chu**[2],
Richard Han[1], **Qin Lv**[1], **Shivakant Mishra**[1]
[1] Dept. of CS, University of Colorado Boulder [2]Microsoft Research
[1] {lei.tian, shaosong.li, junho.ahn, richard.han, qin.lv, mishras}@colorado.edu
[2] davidchu@microsoft.com

ABSTRACT
Online video chat services such as Chatroulette and Omegle randomly match users in video chat sessions and have become increasingly popular, with tens of thousands of users online at anytime during a day. Our interest is in examining user behavior in the growing domain of mobile video, and in particular how users behave in such video chat services as they are extended onto mobile clients. To date, over four thousand people have downloaded and used our Android-based mobile client, which was developed to be compatible with an existing video chat service. The paper provides a first-ever detailed large scale study of mobile user behavior in a random video chat service over a three week period. This study identifies major characteristics such as mobile user session durations, time of use, demographic distribution and the large number of brief sessions that users click through to find good matches. Through content analysis of video and audio, as well as analysis of texting and clicking behavior, we discover key correlations among these characteristics, e.g., normal mobile users are highly correlated with using the front camera and with the presence of a face, whereas misbehaving mobile users have a high negative correlation with the presence of a face.

AUTHOR KEYWORDS
Mobile video chat; behavior analysis; effective matching; misbehavior detection

ACM CLASSIFICATION KEYWORDS
H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems: Video, Evaluation/methodology

INTRODUCTION
Realtime, interactive video-based services are fast becoming an integral part of the Internet user experience. Video chat services from providers such as Skype, Google+ and Facebook are now commonplace. Jump started by Apple's Facetime, these services are even rapidly becoming mainstream for mobile devices such as phones and tablets as well. A common theme is that these video chat services link users who have previously established friend relationships.

On the other hand, *random video chat* services such as Chatroulette [5], Omegle[18] and MeetMe [16] have recently started to gain in popularity. In random video chat, strangers are randomly paired together for video-based conversations. The appeal of these services for users is a classic one: the ability to meet new people face to face, now shifted to the virtual domain. These services have become extremely popular over the last few years. For example, both Chatroulette and Omegle have tens of thousands of users at any given time actively using their systems. As random video chat continues to gain in popularity, we expect *mobile* random video chat to increase in volume and frequency as well.

While standard and even random video chat services have been studied extensively (see, e.g., [11, 1, 24, 6, 25]), prior work has not shed light on the behavior of users in mobile random video chat services. Introducing the new paradigm of mobile interaction into what had formerly been primarily a desktop-based interaction paradigm with webcam-driven online random video chat raises new questions: How are mobile users using random video chat? Is physical mobility fundamentally altering the nature of interacting with a virtual acquaintance?

There are two important motivations for understanding the behavior of mobile users in random video chat. First, from a social science perspective, better insight into why users engage in these services can aid in designs that improve user experience. For example, matching users quickly with other users with whom they would prefer to chat would lead to higher user satisfaction and presumably extended session durations. An understanding of mobile user behavior can inform us as to salient characteristics that would lead to better matching beyond purely random pairings. Second, safeguarding regular users from misbehaving users is equally important. Standard (non-mobile) random video chat services are especially prone to objectionable content such as unauthorized advertisements and sexually explicit flashers who expose all or parts of their bodies. Prior work has shown that intelligent filtering that leverages user behavior patterns can help to safeguard regular users from such content [24, 6,

UbiComp'13, September 8–12, 2013, Zurich, Switzerland.

http://dx.doi.org/10.1145/2493432.2493488

25]. An understanding of misbehavior in the mobile setting can help us extend prior protections to the mobile domain.

However, studying mobile user behavior in random video chat services is challenging due to two compounding factors. First, the very nature of mobility means that users have frequent, untethered and low-overhead opportunities to interact with the video chat service. Therefore, outside observation of active mobile users is a myopic undertaking. Second, as we will show in detail, users are very selective about with whom they interact. Small scale user studies are thus ill-suited to capturing the aggregate trends of random user-to-user interactions, and the formation of new social bonds.

In this paper, we address these important challenges by conducting a large-scale study of mobile-client user behavior in online video chat services in the real world. We have developed *MVChat*, an Android-based mobile client compliant with the popular Omegle random video chat service (Figure 1). The first version was released in November of 2012, and has been used by more than four thousand users so far. In this paper, we analyzed in detail three weeks (January 25th - February 14th) of user behavior on the mobile video chat client.[1] Our key findings are that: the majority of sessions are brief, as users search for a more meaningful partner to converse with for a longer duration session; normal mobile users are highly correlated with using the front camera and with the presence of a face, whereas misbehaving mobile users have a high negative correlation with the presence of a face; females are highly popular, in that users with a large enough fraction of sustained sessions are disproportionately female, but surprisingly gender is not correlated with misbehavior, i.e., females were just as likely to misbehave as males; mobile content is more diverse than Web online content; and groups typically imply normal behavior.

To our knowledge, this is the first-ever study of mobile video chat behavior at scale. In the following, we describe related work, our methodology, the system used for data collection, and our key findings in more detail.

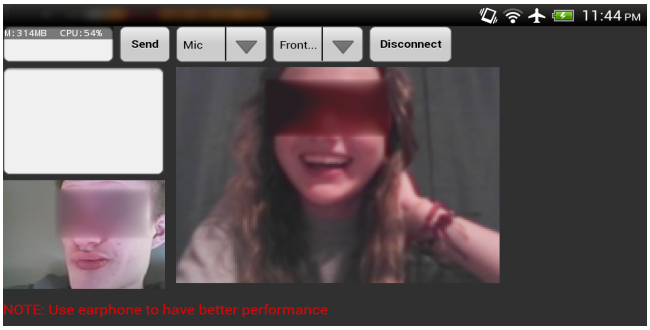

Figure 1. A screenshot of the MVChat application.

RELATED WORK

Prior work by Jana et al. [14] has discussed the issues and challenges in designing successful mobile video chat applications. Scholl et al. [20] designed and built an online video chat application, reporting results from 53 users in a social setting, focusing on bandwidth issues and view navigation. VideoPal explored the use of video to facilitate asynchronous communication between six children and their close friends [11]. Other work focuses on understanding video chat system usage between teenagers [4, 21] and in the context of families, especially the facilitation of communication between grandparents and grandchildren [1, 19, 17]. Our work differs from these prior works in one or more of the following respects: its larger scale; its publicly available deployment; its focus on the mobile context; and the study of random video chat.

Prior research has described the design of classifiers to detect and remove misbehaving users from online random video chat sessions [24, 6, 25]. These papers do not explore the mobile context nor do they provide a detailed understanding of typical and misbehaving user behavior and their correlation factors.

In mobile video research, MoVi explored the use of collaborative sensing on mobile phones to trigger the video recording of a social event by one of the participants' camera phone, as well as the generation of video highlights of the event [2]. MicroCast sought to share video streaming amongst a local group of smartphones, who also share their partial results with one another [15]. Mobile video encoding for wireless links has recently introduced cross-layer encoding (SoftCast) [13], and reliable coding (ChitChat) [23] techniques. A study measuring the energy and bandwidth costs of streaming video from popular websites such as YouTube to six different smartphones has been conducted [10]. More generally, there have been a number of studies examining user behaviors and usage of mobile phones [9, 7, 22, 8, 12, 3].

METHODOLOGY

In seeking to understand user behavior of random mobile video chat at scale, we are interested in answering a variety of questions, starting with demographic questions. Who uses these kinds of services (country of origin, male/female, single/group, etc.)? How well equipped is the software and hardware of smartphones for such users? And which users participate prominently in these services?

Behavioral questions are also of great interest. What is the length of a typical random video chat session? How often do users seek new pairing sessions while using the application? What (day/night/hour) are the most popular times for random video chat? How often do mobile users terminate their sessions compared to the other side terminating the session? What role does texting play during a typical mobile video chat session? What fraction of mobile users behave in an unsafe manner, e.g., flash or reveal themselves? And what are the key differences in user behaviors between mobile users and other online video chat users? Finally, from a contextual standpoint, to what extent can mobile sensor data help us understand the overall environment, e.g., background sound, location (indoor/outdoor), motion, etc. for mobile video chat?

[1]This study has been approved by our institution's Human Research & the Institutional Review Board.

Once we have the basic taxonomic classifications, we ask whether there are any strong correlations (positive or negative) between various characteristics, such as gender, presence/absence of a face, presence/absence of a person, front/back camera usage, normal behavior/misbehavior, and audio silence/voice/music? Is there any correlation between acceleration data and user behavior?

To answer these questions, we have developed an Android-based mobile client (called MVChat) for random online video chat services through which we can collect user data. This client allows mobile users to connect with online Omegle users for random video chat. In addition, MVChat logs on our server multi-dimensional, user-related sensor data from each participating mobile client. This data includes image snapshots, audio, camera position (front or back), accelerometer and gyroscope data, texting data, nexting clicks, and times when new chat sessions begin and end. This data is analyzed and classified on an image, session, and user basis. Image-based analysis examines user behavior based on individual snapshot images and the corresponding sensor data, e.g., audio and accelerometer, nearest in time to when that snapshot was taken. Session-based analysis examines user behavior across all data collected during a single chat session, which contains multiple image snapshots. User-based analysis groups together all data collected during all sessions associated with a single user for behavioral analysis.

Images were categorized depending on whether the snapshot contained one person, more than one person, or no person at all. For images that contain at least one person, we further subdivide them into several different categories: images that contain a face or no face; images that contain a male, female, or mixed; images that contain a normal user or misbehaving user, where a misbehaving user is defined as one that flashes. For audio analysis, we listened to the audio samples and classified them as containing either silence or background noise, human voice, or music & sound. For text analysis, we classified them as either containing some text or not containing any text. Camera position was classified as either front position or back position. Finally, each session was classified as either terminated by the mobile user by clicking the Next button or terminated by the other side.

SYSTEM FOR DATA COLLECTION

In order to understand user behavior at scale, we designed our MVChat system to collect data on the scale of thousands of mobile video chat users, millions of video chat sessions, and gigabytes of image snapshots, audio, and smartphone sensor data. Building a mobile client application that is compatible with an existing online video chat service allows us to quickly scale our study to a large number of users of our application. We chose to make MVChat compatible with Omegle, which is a popular video chat service that has tens of thousands of randomly paired users at any given time and is fairly similar to Chatroulette. Unlike Chatroulette, Omegle's user population is unfiltered, which allows MVChat to gain insight about misbehaving users and thereby capture a more representative sample of the true proportions and behaviors of mobile video chat users.

We chose the Android mobile development platform because the application approval process for iPhone essentially precludes any applications such as random video chat that have some indecent content or misbehaving users. Our system consists of three major components: an Android mobile video chat client that is Omegle compliant, a data collection server to store multi-dimensional mobile user data, and the Omegle server. Figure 2 illustrates the overall architecture of the MVChat system.

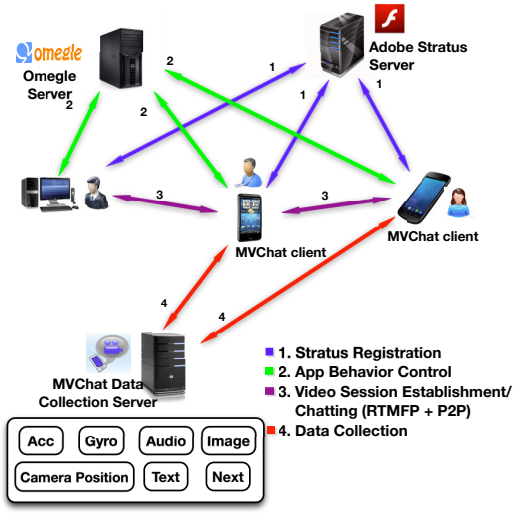

Figure 2. System architecture of the MVChat system.

Compatibility of the MVChat clients with the Omegle system required careful engineering to mimic the behavior of online Web clients. To start an Omegle video chat session, a typical Web browser/client contacts the Adobe stratus server to register and receive a unique peer ID. Next, that client establishes a connection with the Omegle server by providing it the peer ID. The connections between the Omegle server and all the Web clients are used for establishing random pairing of online Omegle clients for chatting. Each time a client requests to chat with someone, the Omegle server replies with the peer ID of a randomly chosen client that is currently idle. The Omegle server also notifies the randomly chosen client with the requesting client's peer ID. The two clients then proceed to establish a peer-to-peer video session between them using each other's peer IDs and the Real Time Media Flow Protocol (RTMFP).

After the video session has been established, each MVChat mobile client establishes a separate connection to our data collection server and periodically posts to it the user's image, audio, camera position, accelerometer and gyroscope data as well as any text data. When a client wants to end its video chat session, it notifies the Omegle server and at the same time stops transmitting its video stream, releases resources, and stops posting data to the data collection server.

Mobile Client Application

Figure 1 shows a screenshot of the MVChat mobile client. Unlike the Omegle Web client, where text messaging dominates the screen and the videos of the sender and receiver are shown in less than half the screen, our MVChat mobile client

emphasized video first, as there was limited screen real estate and we felt most mobile users would interact most easily with video and audio. After invocation and connection with a remote user, the mobile client displays the remote client's video (captured via remote user's device camera) in the large window on the right. The local camera view that is being sent to the remote user is shown in the lower left window. However, to maintain compatibility with the user experience of Omegle chatters, we included text messaging. Text messages are composed in the upper left dialog box and the most recent messages are shown in the middle left box.

To maintain compatibility with Omegle's session flow, where one click ends a session and a second click requests a new session, we added a Disconnect button that is displayed at the top when a video chat session is in progress. A user may press this button to end the current video chat session, which changes the button to a Next button, and then a user may press the Next button to request a new client to chat with. Unlike Web clients, mobile devices have both front and back cameras, so we added a pull-down menu that allows the user to specify whether they wish to use their front or back camera for a chat, and this can be changed anytime during a video chat session. The other pull-down menu allows the user to specify the microphone that they wish to use, should there be more than one supported.

Data Collection Server

The data collection server is comprised of three components: an Apache server, a MySQL server and a Flash Interactive Media Server. The Apache server stores image snapshots directly on our server file system and cooperates with the MySQL server to store most of the rest of the mobile sensor data. The Flash Interactive Media Server is responsible for storing the audio file and for easy binding with the mobile client via the RTMFP protocol.

For each mobile device, a random device ID (not the Adobe Stratus peer ID) is generated, and for each video session, the device generates a session ID. Every set of data posted from devices to the data collection server is tagged with the captured time stamp, camera position as well as the device and session IDs for easy segmentation.

The default sampling frequency for accelerometer and gyroscope is 5Hz. Snapshots are captured every 30 seconds and audio is captured in the first 10 seconds of every 40-second interval. On average, each image is about 35 KB (120*160) and each audio file (10 second duration) is around 110 KB. So the total amount of data transmitted from mobile client to the data collection server is less than 5 KB per second per user. This aggregate sampling rate in theory allows us to scale to thousands of concurrent users, though in practice we have thus far seen a maximum of 80 concurrent users.

In order to accommodate unforeseen workloads and help the system scale, we designed the data collection server to adjust flexibly or even turn off the sensor reporting streams from the mobile clients. For each mobile client, the server can vary the sampling frequencies, or even stop data collection altogether. This ability allows us to throttle clients if the overall load on the data collection server is too high, and also to suspend data collection while allowing the clients to continue to video chat normally. When the client starts, it first contacts the data collection server, which responds with one of three options: disabling the client completely for this session so that it cannot access the Omegle service at all; disabling just the posting of sensor data to the collection server, whereupon the client can access the Omegle service without any remote logging; or enabling the posting of sensor data as the user accesses the Omegle service. In the third case, the server also (optionally) specifies the sampling frequencies for each sensor modality.

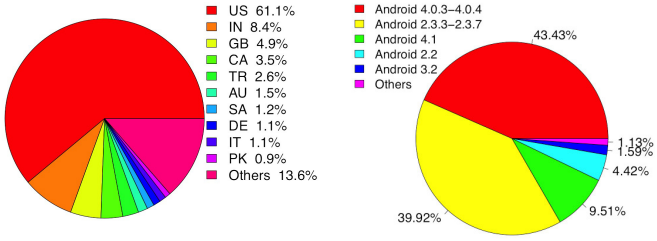

Figure 3. (a) User country distribution and (b) installation device OS distribution.

Data Collection Experiment

We deployed our MVChat mobile application on the Google Play marketplace for Android applications and collected data spanning about 3 weeks from January 25th to February 14th, 2013. In total, 4,632 distinct users of our MVChat application were identified, generating 1,703,837 pairing sessions. To protect remote users' privacy, data was only collected for our mobile users, and not for the remote users paired with them. The total amount of data collected was about 170 GB, comprised of 70 GB of image snapshots, 90 GB of audio snippets, 8.5 GB of mobile sensor, texting, and clicking data.

Figure 3(a) shows the distribution of users by country of origin who have downloaded and installed our application. We see that the vast majority are from the United States and that all other countries comprise at most single digit percentages of the downloads. As shown in Figure 3(b), many users of our application still rely on older versions of Android 2.3, but that a fairly large fraction have Android 4.0 or higher. We also found that about three-fourths of our users' smartphones were equipped with both front and back cameras, but that a fourth still lacked the front camera.

USER BEHAVIOR ANALYSIS

We conducted a comprehensive analysis of our data to understand and identify key behavioral characteristics of mobile video chat users at scale. Our data analysis consists of five specific components: (1) *Global data analysis* studies the overall statistical distributions and identifies common user behaviors. (2) *Mobile-online video chat comparison* highlights the differences between mobile and online video chat, and new features associated with mobile video chat. (3) *Taxonomy correlation analysis* identifies user behavioral attributes that have strong positive or negative correlations

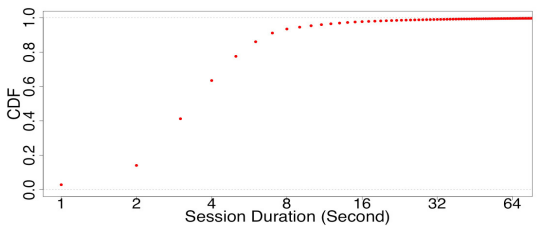

Figure 4. CDF of session duration: 99.5% of the 1.7 million video chat sessions were shorter than 60 seconds.

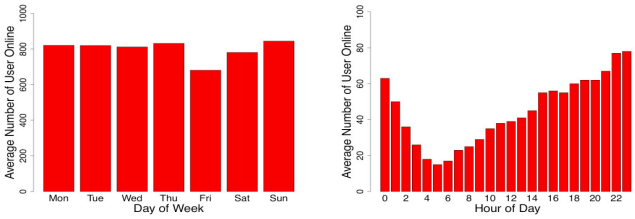

Figure 6. Time of use: (L) day of week and (R) hour of day.

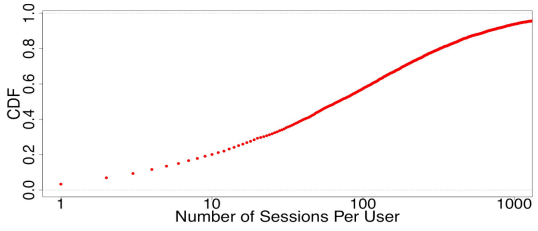

Figure 5. CDF of number of sessions per user: 42% of the 4,632 users had more than 100 video chat sessions.

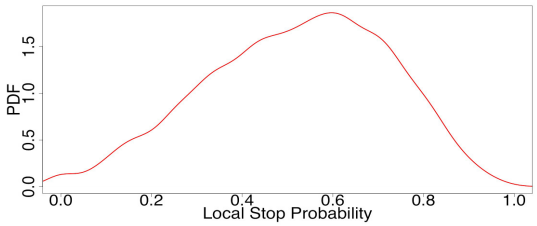

Figure 7. PDF of local stop probability: Mobile users are more likely to end video chat sessions locally.

and directional association rules with strong confidence. (4) *Meaningful user analysis* aims to characterize users who are popular and are more effective in maintaining longer video chat sessions. (5) *Accelerometer data analysis* studies how accelerometer sensor data can help identify certain user behavioral characteristics in mobile video chat.

Global Data Analysis

We first conduct a global data analysis to understand key characteristics of user behavior in mobile video chat, including session duration, time of use, local stop behavior, use of text messaging, and a taxonomy distribution.

Session duration refers to the length of users' video chat sessions. Figure 4 shows the cumulative distribution function (CDF) for session duration. Among the 1.7 million video chat sessions we have collected, 80% of the sessions were less than 5 seconds; only 1% of the sessions were longer than 30 seconds; and only 0.5% of the sessions were longer than 60 seconds. Figure 5 shows the CDF of number of sessions for each user. Among the 4,632 users we have observed, 80% of the users participated in at least 10 sessions, 42% of the users participated in more than 100 video chat sessions; and 6% of the users participated in over 1,000 video chat sessions during our 3-week data collection period. A hypothesis consistent with these findings is that video chat users spend a lot of effort going through many random pairing sessions in order to find someone interesting to chat with for a longer duration. This suggests that random pairing is ineffective in the sense of generating a lot of "noise" until the "right" pairing shows up in a video chat screen. While some video chat services have recently introduced interest-based pairing, it is not clear if interest is the only and most effective matching metric. To better understand this problem, we study "meaningful sessions" and "meaningful users" later in this section.

Time of use refers to the time when users participate in video chat sessions. We consider both day of week and hour of day using the local time reported by users' smartphones. As shown in Figure 6(L), the average number of users is similar throughout a week except for Fridays and Saturdays. This is interesting as we expect that more users have free time on Fridays and Saturdays. One possible explanation is that, because more users are free on Fridays and Saturdays, they go out and participate in other social events (i.e., "party time") instead of using our mobile video chat application. Figure 6(R) shows the number of users during different hours of the day. We can see that early morning (5am) has the fewest number of users, and the average number of users increases steadily throughout the day, with quick jumps at 3pm and 10pm, and reaches the highest number around midnight. The fact that we observe more users during the late evening hours is potentially related to the private nature of mobile video chat, i.e., when users are by themselves and would like to meet strangers in the virtual world.

Local stop probability refers to the probability of a user ending his/her session locally instead of the remote party ending the session. Note that each user can click the "Disconnect" and "Next" buttons at any time to end the current video chat session and start a new pairing session. If we assume that for each video chat session, the two parties (local and remote) have equal probability of ending the session, then across all users and their sessions, we would expect to see a distribution for local stop probability with a mean value of 0.5. However, the distribution shown in Figure 7 is different, with a mean value close to 0.6. Note that we only collected data from mobile video chat users and not the online video chat users at Omegle with whom our mobile users chat. The higher local stop probability of our mobile users means that mobile users are more likely to end a video chat session than the online users. We think such deviation is possibly due to the difference in network connection between mobile users and online users. Video chat is relatively expensive and requires good network connection. Our mobile users are typically connected through 3G or WiFi networks, while the online users usually have better network connection with higher bandwidth and more stability. As a result, mobile users are more likely to end a session if there is delay in starting the session or the quality is low.

Table 1. Taxonomy Distribution of Meaningful Session Images

Taxonomy	Distribution			
Person	No Person	Person		
	594	1519		
Group	No Person	Single	Group	
	594	1446	73	
Normal	No Person	Misbehaving	Normal	
	594	314	1205	
Face	No Person	No Face	Face	
	594	425	1094	
Gender*	No Person	Male	Female	Both
	594	688	721	3
Camera Position	Back	Front		
	653	1460		
Audio	No Data	Silence/Noise	Voice	Music
	687	735	594	94

*There are 107 images for which we could not determine the gender.

Table 2. Taxonomy Distribution of Meaningful Sessions

Taxonomy	Distribution			
Person	No Person	Person		
	62	156		
Group	No Person	Single	Group	
	62	147	9	
Normal	No Person	Misbehaving	Normal	
	62	57	99	
Face	No Person	No Face	Face	
	62	50	106	
Gender	No Person	Male	Female	Both
	62	94	50	1
Text	No Text	Text		
	90	128		
Stop	No Data	Local Stop	Remote Stop	
	49	79	90	
Audio	No Data	Silence&Noise	Voice	Music
	4	122	79	13

Text messages can also be used in users' video chat sessions. Although 58.3% of the users had used text messages during video chat, only 2.8% of all video chat sessions contained text messages. We think this is due to the fact that most of the sessions were short sessions, when users are quickly clicking through many random pairing sessions in order to find the right person to talk to. These short sessions generally contained no text.

GPS data collection is a function incorporated in our application. However, the data we collected contain almost no GPS data. This, together with our examination of the snapshot images, indicates that almost all video chat sessions occurred indoors. This is reasonable as video chatting with strangers is considered a private activity and people prefer to participate in such activities in private indoor environments. Moreover, based on our observation of the snapshot images, many of our mobile video chat users are young people and tend to use the application in their homes or dorms. The indoor locations vary from living room to bedroom and even bathroom. Also, mobile user posture is distributed across sitting, lying down, and standing, and appears to be more diverse than online Webcam-based images captured from desktop clients [24], where users typically are located in the bedroom and are sitting.

Taxonomy analysis aims to characterize key user behaviors when using the mobile video chat application. Due to the large scale of the data we have collected (4,632 users and 1.7 million sessions), it is infeasible to label all the data. In addition, we have observed earlier that the majority of the sessions were short sessions. Therefore, we decided to focus our taxonomy analysis on *meaningful sessions* sampled from the overall data set, i.e., sessions that lasted 60 seconds or longer. Our reasons for focusing on meaningful sessions are three-fold: (1) We want to understand what user behavioral characteristics promote longer and potentially more effective video conversations; (2) Sessions that last at least 60 seconds contain at least 3 snapshot images and 2 audio samples, which provide adequate information to label each session; and (3) There are almost 8,000 meaningful sessions in our data set, which are sufficient for our analysis. Among all the meaningful sessions, we randomly sampled 1/40 of the sessions. After removing noisy sessions whose snapshots are black and have no content, we have a set of 218 meaningful sessions with 2,113 images in total. Using the taxonomy we have defined, we then manually label each individual image. Since each meaningful session contains multiple snapshot images and the variance of these images' labels is generally small, we use majority voting of image-based labels to determine the taxonomy labels for each session. There are two exceptions. First, camera position has a higher variance in sessions, so we ignore them in session-based analysis. Second, to label normal and misbehaving sessions – if any image in a session is labeled as misbehaving, that the whole session is labeled as misbehaving. For each session, we also considered text and local vs. remote stop information.

The taxonomy distributions of meaningful session images and meaningful sessions are summarized in Table 1 and Table 2, respectively. We can see that most meaningful sessions and their images contained a person, single user, normal user, user face, silence/noise or voice. Also, the front camera is used much more often than the back camera. It is interesting to note that among the meaningful sessions, there are fewer female sessions than male sessions, but more female images than male images. The reason is that female users are possibly more popular and tend to have longer video chat sessions, thus each session contains more images.

Mobile-Online Video Chat Comparison

We compared both mobile and online user behavior in the Omegle application, with the caveat that the information we obtained from online users was less extensive than our mobile data collection. This is because we relied on a randomized image data set provided to us by Omegle, from which we could derive such properties as gender proportion, but not session duration for example. While we could have built our own instrumented Web application compliant with Omegle to measure these factors, we felt the adoption rate would not have been strong since Web browsers already connect with Omegle. In contrast, our mobile application introduced new capabilities - namely mobile video chat - that drove adoption. Figure 8 compares the gender distributions of mobile and online Omegle users. As shown in the figure, mobile users were about 49% male and 12% female, while online users were 68% male and 17% female. Note mobile users had a much higher fraction of other types of content. Due to the portable feature of mobile video chat, users can engage

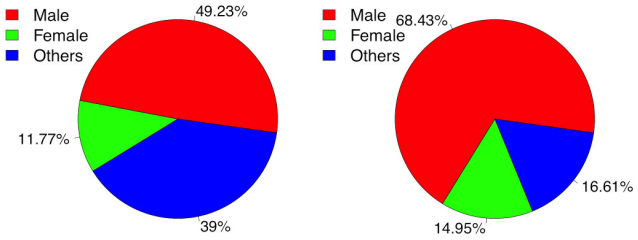

Figure 8. Comparison of user gender distribution: (L) mobile video chat and (R) online video chat. "Others" refer to other types of content.

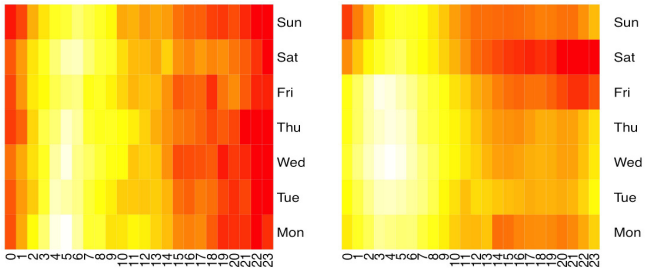

Figure 10. Comparison of time of use (day of week and hour of day): (L) mobile video chat and (R) online video chat.

in video chat sessions at varying locations and with different posture. Many smartphones are also equipped with both front and back cameras, allowing users to switch between the two cameras and show different content during video chat sessions. As a result, with mobile video chat, we expect to see more diverse other types of content than that of online video chat, such as the sample images shown in Figure 9.

We also compare the time of use, i.e., day of week and hour of day, between mobile and online Web users. The results are shown in Figure 10. Omegle is a popular online video chat service, and its number of concurrent users vary between 15,000 and 45,000. For our application, the number of users at any given hour of day and given day of week varies between 15 and 80. The general trend for hour of day is similar for both mobile and online users. Overall, online users are more uniformly distributed across different time, while mobile usage sees more fluctuation (partly due to the smaller number of mobile users) and higher usage in the afternoon and late evening hours. One significant difference is on day of week: online usage is much higher on Friday evenings and most of Saturdays, while mobile usage actually sees lower activity during these two days. This may be due to different user populations for the mobile and online worlds, e.g., mobile users are more likely to be out partying on Fridays and Saturdays.

Taxonomy Correlation Analysis

Our taxonomy aims to characterize users' mobile video chat behavior from multiple dimensions. Given the labeled taxonomy information, one important question we want to answer is *which user characteristics are correlated*. In other words, we want to identify behavioral characteristics which are likely or unlikely to occur together. For instance, do male and female users behave differently, or how do normal users behave compared to misbehaving users. Understanding such taxonomy correlations can offer useful insights into designing better user pairing strategies, misbehavior detection mechanisms, etc.

Since our taxonomy contains categorical rather than numerical attribute values, we utilize four correlation metrics that are typically used for categorical correlation analysis: χ^2, $lift$, $all\_confidence$ (or all_conf), and $cosine$. Let A and B be two attributes (e.g., gender and camera position) with values $a_i (1 \leq i \leq c)$ and $b_j (1 \leq j \leq r)$ respectively, these four metrics are defined as follows:

$$\chi^2 = \sum_{i=1}^{c}\sum_{j=1}^{r}\frac{(n_{ij}-e_{ij})^2}{e_{ij}} \quad (1)$$

$$e_{ij} = \frac{count(A=a_i)*count(B=b_j)}{N} \quad (2)$$

$$\text{lift}_{ij} = n_{ij}/e_{ij} \quad (3)$$

$$\text{all\_conf}_{ij} = \frac{n_{ij}}{\max\{count(A=a_i), count(B=b_j)\}} \quad (4)$$

$$\text{cosine}_{ij} = \frac{n_{ij}}{\sqrt{count(A=a_i)*count(B=b_j)}} \quad (5)$$

Here N is the total number of samples, n_{ij} is the number of samples with both $A=a_i$ and $B=b_j$, $count(A=a_i)$ and $count(B=b_j)$ are the numbers of samples with $A=a_i$ and $B=b_j$ respectively.

χ^2 measures the difference between observed values n_{ij} and expected values e_{ij} (if A and B are not correlated). So a small χ^2 value (close to 0) means non-correlation while a high χ^2 value indicates possible correlation. Similarly, a $lift$ value of 1 means no correlation ($n_{ij}=e_{ij}$), and a $lift$ value >1 (or <1) indicates positive (or negative) correlation. However, both χ^2 and $lift$ are sensitive to skewed distribution of the attribute values (e.g., most user are normal or very few users chat as a group). By focusing on a_i and b_j values and ignoring other values (i.e., null-invariant), all_conf and $cosine$ can tolerate different data set scales and skewed attribute value distributions. Generally, all_conf and $cosine$ values that are close to 1 indicate strong positive correlation and values that are close to 0 indicate strong negative correlation. In our analysis, we leverage χ^2 and $lift$ to confirm non-correlation (i.e., χ^2 is close to 0 and $lift$ is close to 1), and leverage all_conf and $cosine$ to identify strong positive correlation (i.e., both all_conf and $cosine$ are >0.85) and strong negative correlation (i.e., both all_conf and $cosine$ are <0.1).

Note that correlation is bi-directional: if a_i and b_j are positively (negatively) correlated, seeing a_i means that b_j is more (less) likely to occur, and vice versa. However, some relations between user behaviors can be unidirectional, e.g., if a_i occurs, b_j is likely to occur but the reverse may not be true. We use association rules (e.g., $a_i \Rightarrow b_j$) to capture such one directional relations, and the conditional probability $Pr(b_j|a_i)$ is referred to as the *confidence* of an association rule.

Figure 11 summarizes the key results of our taxonomy correlation analysis, including strong positive (negative) correla-

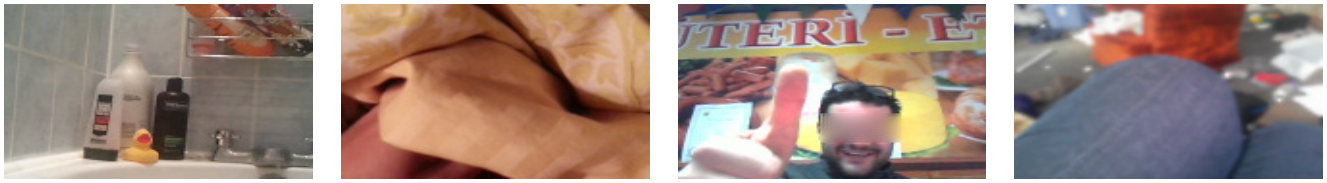

Figure 9. Sample images collected via our app demonstrating more diverse image content in mobile video chat than that in online video chat.

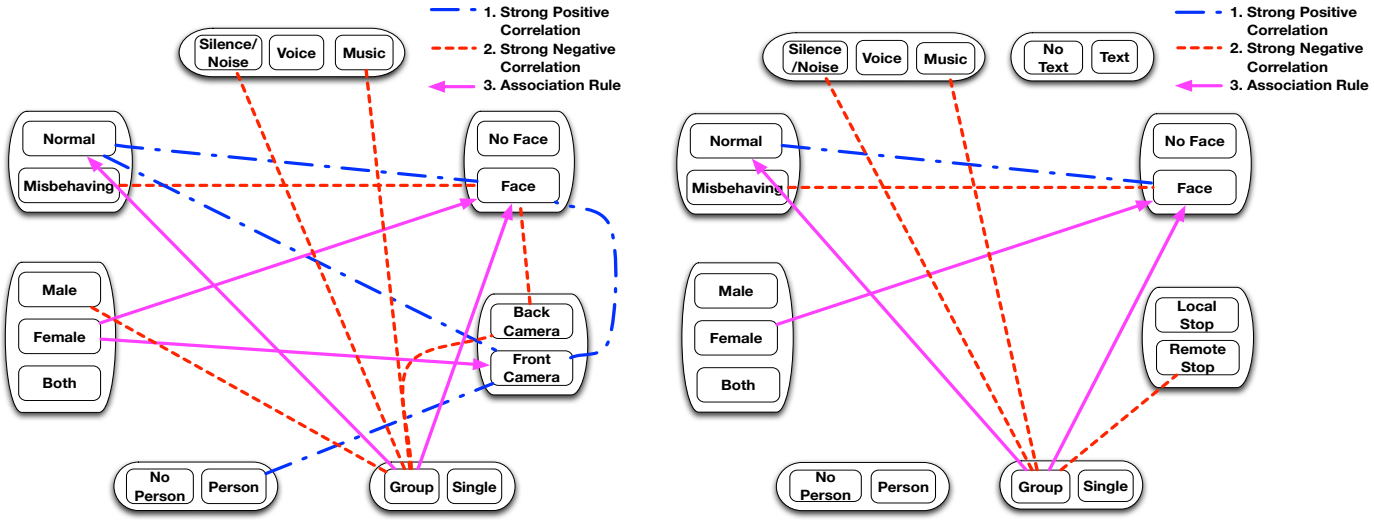

Figure 11. Taxonomy correlation analysis: (L) image-based graph and (R) session-based graph. Highlighted in the graphs are strong positive correlations (both all_conf and $cosine$ are > 0.85), strong negative correlations (both all_conf and $cosine$ are < 0.1), some non-correlations (χ^2 is close to 0 and $lift$ is close to 1), and association rules with $confidence > 0.8$.

tions, some (surprising) non-correlations, and one-directional association rules with high confidence values. Our correlation analysis has been conducted at both the image level and session level. Image-based analysis broadly refers to the multi-modal set of data closest in time to an image snapshot, including the audio snippet and sensor readings immediately preceding a snapshot. Note that our sampled meaningful session data set contains 218 sessions and 2,113 images, so the image-based analysis is more stable, but the session-based analysis still offers some important insights. Next, we describe in detail the image-based correlation analysis results, then discuss the differences in the session-based correlation analysis results.

Camera position, i.e., front or back camera, is a key feature that allows users to show different content. As shown in Figure 11(L), *Front Camera* has strong positive correlations with *Person* (seeing person in the image), *Face* (seeing face in the image), and *Normal* (normal user), and strong negative correlation with *No Face* (not seeing face in the image); while *Back Camera* has strong positive correlation with *No Face* and strong negative correlation with *Face*. This can be explained by the notion that people typically use the front camera to show their own faces and the back camera to show some other content. In addition, the observation that normal users and front camera have a strong positive correlation can be used to differentiate normal users from misbehaving ones. Since checking camera position on smartphones is an inexpensive operation, this can be particularly useful for misbehavior detection in mobile video chat services.

Face appearance in video chat is another important factor to consider. In video chat services, people seek other interesting people to chat with, and showing their faces help keep people engaged in video chat sessions. As shown in Figure 11(L), *Face* has a strong positive correlation with *Normal* and a strong negative correlation with *Misbehaving*. In other words, normal users tend to show their faces while misbehaving users tend not to show their faces. We believe the explanation is three-fold: (1) Normal users show their faces so they can chat more effectively with their partners; (2) Misbehaving users tend to hide their faces to avoid being identified; and (3) Due to the limited aperture angle of mobile phone cameras, it is difficult for a misbehavior to show both his/her face and private body parts. Given the strong positive (negative) correlations between Face and Normal (Misbehaving), image-based face detection can be quite effective for differentiating normal users from misbehaving ones [24, 25].

Group chatting is rare in video chat services and most users choose to chat alone with their remote partners (Table 1, Table 2). Still, when group chatting does occur, we observe that *Group* has strong negative correlations with *Back Camera*, *Music & Sound*, *Silence/Noise*, and *Male*. In other words, when people chat as a group, they are less likely to use back camera, have background music/sound or silence/noise. It is

also interesting to observe that male users tend not to chat in groups. In addition, based on the association rules shown in Figure 11(L), when users chat as a group, they are very likely to show their faces and are very likely to be normal.

Female users also have some interesting behavioral characteristics: They are likely to use the front camera and show their faces (Figure 11(L)). Specifically, based on our sampled data set, the probability for a female user to use the front camera is 92% and the probability for a female user to show her face is 84%. The intuition is that female users tend to be popular in video chat services, i.e., more people are interested in talking to female users. Therefore, using the front camera and showing their faces can help female users to chat more effectively with their remote partners.

Non-correlations can sometimes indicate something interesting as well. For instance, we observe no negative correlation between *No Person* and *Voice*, i.e., even when there is no person shown in the video chat images, there can be human voice in the audio recordings. This may indicate that users sometimes show other content (not themselves) to their remote partners while talking. Another example is Gender versus Normal/Misbehaving. We originally expected to see *Male* being correlated with *Misbehaving*, i.e., male users are more likely to misbehave and misbehaving users are more likely to be male. However, we did not find such correlation in our data set. In other words, a misbehaving user can be male or female, and knowing the gender of a user does not increase or decrease his/her probability of misbehaving.

Session-based correlation analysis reveals fewer relations than image-based correlation analysis (Figure 11). This is mainly due to the fact that there are fewer number of sessions than images. Most of the session-based relations are similar to that of image-based relations. One correlation that is specific to session-based analysis is the strong negative correlation between *Group* and *Remote Stop*, i.e., video chat sessions of group users are unlikely to be stopped by the remote party. The fact that people do not hang up when talking to a group of users on the remote side is interesting. Although group chatting is rare in our current data set, they do seem to keep people more engaged in a video chat session.

Meaningful User Analysis

Besides understanding the characteristics of meaningful sessions ($>$ 60 seconds), we also want to characterize "meaningful users", i.e., users who are more successful in participating in longer video chat sessions. We define meaningful users as the users who have more than 10% probability to have a video chat session that is more than 30 seconds long. Given that most users have less than 5% probability to have longer than 30 seconds sessions, 10% is considered significant. We pick 30 seconds instead of 60 seconds as the session duration threshold, since the former allows us to study 114 meaningful users while the latter has only 30 meaningful users. For comparison purposes, we also sampled a similar number of "non-meaningful users". And for each meaningful or non-meaningful user, we randomly sampled about 20 images from the user's sessions. As a result,

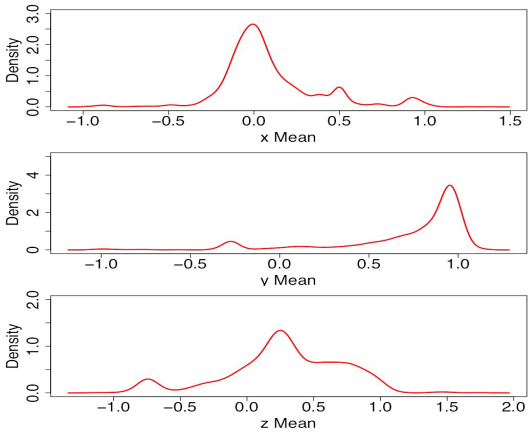

Figure 12. Acceleration distribution of meaningful sessions.

we have obtained 2,714 images for 114 meaningful users and 2,832 images for 123 non-meaningful users. We labeled these images using the same taxonomy as defined before. By comparing the taxonomy distributions of these two different user sets, we find that *gender* is the most dominating factor. In the set of meaningful users, we have 10 males and 69 females (12.7% vs. 87.3%), while in the set of non-meaningful users, we have 56 males and 9 females (86.2% vs. 13.8%). Note that the remaining 35 'meaningful users' and 58 'non-meaningful users" images are absent of person content so we cannot identify their genders. A similar gender distribution occurs when using the 60 second definition, where the skew is even more significant: 2 males and 20 females (9.1% vs. 90.9%). To summarize, compared with male users, female users are much more likely to have long meaningful sessions. This is partially due to the fact that there are a lot more male users than female users in video chat services (Figure 8), and male users appear to be more interested in talking with female users.

Acceleration Analysis

For mobile video chat, acceleration information can be easily collected and can offer useful contextual information such as how the phones are positioned/oriented during video chat. Figure 12 shows the acceleration distribution of meaningful sessions. Note that the Adobe AIR API reverses the x and y axes used in the Android API, and refers to the long side of the mobile phone panel as the x axis (and the short side as the y axis). Since MVChat organizes its video chat screen using the landscape layout, users tend to hold the phone horizontally and potentially rotate along the x axis. As a result, acceleration centers around 0 for the x axis, centers around G for the y axis, and spreads between $-0.5G$ and G for the z axis. Based on the acceleration distribution, we define four different classes to identify acceleration orientation based on the degree of rotation along the x axis (Figure 13). In other words, the phone can be placed horizontally facing up, can be rotated along the long-side phone panel (tilting) until facing down. We found that (1) Class 1 and Class 2 occur much more frequently than the other two classes, (2) normal users are more likely to use the front camera with orientation classes 1 and 2, (3) misbehaving users are more likely to use the back camera with acceleration orientation classes 1 and 2. This can be potentially leveraged to design

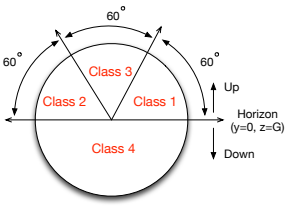

Figure 13. Acceleration orientation classes defined based on the degree of rotation along the x axis (long side plane of smartphones).

more effective misbehavior detection mechanisms for mobile video chat services.

CONCLUSIONS & FUTURE WORK

In this paper, we have presented the results of a large scale study of mobile users of an Android video chat application, which we designed and deployed for random video chat with Omegle online users. We summarize our findings as follows. We found that most sessions were short as users sought interesting people to talk to, which motivates future work for better user matching strategies. Normal users are highly correlated with using the front camera and showing their faces, whereas misbehaving users tend to hide their faces – which suggests the exploration of camera position and face detection for distinguishing normal users from misbehaving ones. Users with a large enough fraction of sustained sessions are disproportionately female, but surprisingly females were just as likely to misbehave as males. Mobility introduces more diversity than Web online content, while groups typically imply normal behavior.

In the future, we plan to increase the scale of our dataset further, incorporate analysis of more sensing modalities such as the gyroscope and text messaging context, and better understand what factors characterize meaningful user behavior, with a view towards potentially more effective matching. We also hope to use our findings to design an accurate misbehavior classifier for mobile video chat users. Our analysis reveals some good indicators for that issue like group users will not misbehave, and different camera positions and orientations inferred by acceleration result in different distributions for normal and misbehaving users.

ACKNOWLEDGMENTS

We would like to acknowledge the support of NSF grants CSR-1162614 and CIC-1048298. We would like to thank Leif K-Brooks, the CEO of Omegle, for helping to enable our MVChat client to be compliant with Omegle's server, and in providing data for our research. We would also like to thank Xinyu Xing, Hanqiang Chen and Yuli Liang for their advice on image analysis. We also sincerely thank the anonymous reviewers for their comments and suggestions.

REFERENCES

1. M. G. Ames, J. Go, J. J. Kaye, and M. Spasojevic. Making love in the network closet: the benefits and work of family videochat. In *Proc of the 2010 ACM conf. on Computer supported cooperative work*, CSCW '10, pages 145–154, 2010.
2. X. Bao and R. Roy Choudhury. Movi: mobile phone based video highlights via collaborative sensing. In *Proc. of 8th Intl. ACM Conf. on Mobile Systems, Applications, and Services*, MobiSys '10, pages 357–370, 2010.
3. M. Böhmer, B. Hecht, J. Schöning, A. Krüger, and G. Bauer. Falling asleep with angry birds, Facebook and kindle: a large scale study on mobile application usage. In *Proc. of the 13th Intl. Conf. on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '11, pages 47–56, 2011.
4. T. Buhler, C. Neustaedter, and S. Hillman. How and why teenagers use video chat. In *Proc. of the 2013 conf. on Computer supported cooperative work*, CSCW '13, pages 759–768, 2013.
5. Chatroulette web site. http://www.chatroulette.com/.
6. H. Cheng, Y.-L. Liang, X. Xing, X. Liu, R. Han, Q. Lv, and S. Mishra. Efficient misbehaving user detection in online video chat services. In *WSDM '12*, pages 23–32, 2012.
7. K. Church and B. Smyth. Understanding mobile information needs. In *Proc. of the 10th intl. conf. on Human computer interaction with mobile devices and services*, MobileHCI '08, pages 493–494, 2008.
8. Y. Cui and V. Roto. How people use the web on mobile devices. In *Proc. of the 17th intl. conf. on World Wide Web*, WWW '08, pages 905–914, 2008.
9. J. Froehlich, M. Y. Chen, S. Consolvo, B. Harrison, and J. A. Landay. Myexperience: a system for in situ tracing and capturing of user feedback on mobile phones. In *Proc. of the 5th intl. conf. on Mobile systems, applications and services*, MobiSys '07, pages 57–70, 2007.
10. M. A. Hoque, M. Siekkinen, J. K. Nurminen, and M. Aalto. Investigating streaming techniques and energy efficiency of mobile video services. Technical report, 2012.
11. K. Inkpen, H. Du, A. Roseway, A. Hoff, and P. Johns. Video kids: augmenting close friendships with asynchronous video conversations in videopal. CHI '12, 2012.
12. http://blog.appsfire.com/infographic-ios-apps-vs-web-apps/.
13. S. Jakubczak and D. Katabi. A cross-layer design for scalable mobile video. In *Proc. of the 17th annual intl. conf. on Mobile computing and networking*, MobiCom '11, pages 289–300, 2011.
14. S. Jana, A. Pande, A. Chan, and P. Mohapatra. Mobile video chat : Issues and challenges. *IEEE Communications Magazine, Consumer Communications and Networking Series*, 2013.
15. L. Keller, A. Le, B. Cici, H. Seferoglu, C. Fragouli, and A. Markopoulou. Microcast: cooperative video streaming on smartphones. In *Proc. of the 8th intl. conf. on Mobile systems, applications, and services*, MobiSys '10, pages 57–70, 2012.
16. MeetMe web site (formerly MyYearbook). http://www.meetme.com/.
17. M. Milliken, S. ODonnell, K. Gibson, and B. Daniels. Older adults and video communications: A case study. *The Journal of Community Informatics*, 8(1), 2012.
18. Omegle web site. http://www.omegle.com/.
19. H. Raffle, G. Revelle, K. Mori, R. Ballagas, K. Buza, H. Horii, J. Kaye, K. Cook, N. Freed, J. Go, and M. Spasojevic. Hello, is grandma there? let's read! storyvisit: family video chat and connected e-books. CHI '11, pages 1195–1204, 2011.
20. J. Scholl, P. Parnes, J. D. McCarthy, and A. Sasse. Designing a large-scale video chat application. In *Proc. of the 13th ACM intl. conf. on Multimedia*, MULTIMEDIA '05, pages 71–80, 2005.
21. Study: 37% of U.S. teens now use video chat, 27% upload videos. http://techcrunch.com/2012/05/03/study-37-of-u-s-teens-now-use-video-chat-27-upload-videos/.
22. H. Verkasalo. Contextual patterns in mobile service usage. *Personal Ubiquitous Comput.*, 13:331–342, 2009.
23. J. Wang. Chitchat: Making video chat robust to packet loss. Master's thesis, Massachusetts Institute of Technology, 2010.
24. X. Xing, Y.-L. Liang, H. Cheng, J. Dang, S. Huang, R. Han, X. Liu, Q. Lv, and S. Mishra. Safevchat: Detecting obscene content and misbehaving users in online video chat services. In *Proc. of the 20th intl. conf. on World Wide Web*, WWW '11, pages 685–694, 2011.
25. X. Xing, Y.-l. Liang, S. Huang, H. Cheng, R. Han, Q. Lv, X. Liu, S. Mishra, and Y. Zhu. Scalable misbehavior detection in online video chat services. In *Proc. of the 18th ACM SIGKDD conf. on Knowledge discoery and data mining*, KDD '12, pages 552–560, 2012.