Quantifying Presence using Calling Patterns

Husain Husna, Santi Phithakkitnukoon, Enkh-Amgalan Baatarjav, and Ram Dantu
Network Security Laboratory
Department of Computer Science and Engineering
University of North Texas
Denton, TX 76203 USA.
{*hjh0036, santi, eb0050, rdantu*}@*unt.edu*

Abstract—**Presence technology is going to be an integral part of the next generation of communication technology. It can eliminate telephone tag between two parties (caller and callee), which will increase productivity of the parties and reduce unnecessary bandwidth usage of unwanted calls. In this paper, we propose a Willingness Estimator that computes willingness level of a callee for receiving calls from a specified caller. By knowing the willingness value of the callee the caller can decide on proceeding with the call or not. The proposed Willingness Estimator is tested with real mobile user data, and these results are highly accurate. We measure willingness based on calling patterns (arrival time, location, day) of a caller to a callee and this could serve as one of the future presence based services. The results can be used in telecommunication networks such as PSTN, Cellular networks, and Voice over IP networks**

I. Introduction

Presence-aware technology makes communication between parties facilitates reduced cost and time by allowing information such as who and when someone is available in a corporate network .According to [1], presence propagates information about a users willingness, ability, and desire to communicate using a variety of mediums. A caller can subscribe to learn callees behavior towards accepting the calls. The result is precise communication that eliminates the inefficiency of phone tags. We already see this kind of improved communication in Instant Messenger (IM) users. IM users can directly communicate to people who are available at present without guessing. The presence in IM is just a beginning of presence-aware technology evolution. The Gartner research [2] predicted that by 2009, 80 percent of business applications will have presence-aware functionality to support business processing and management of customer relationship and corporate performance.

It is evident that the presence service is going to be an important feature of communication systems. This service can provide advantages to both parties. The service can reduce disrupting calls for the callee when he/she is in the middle of any important work. Caller will be able to know how willing the callee is to receive the call. Integrating presence service with SIP protocol can provide multiple medium of communication between parties.

Presence-aware technology has received a lot of attention from the researchers[3][4]. Shan and Shrirams work [5] was to reduce enterprise server load by mobile clients sharing presence information within a network, and only one of the client acts as a gateway to interact with the server to supply the presence information of the network.

Real-life traffic profile: In this paper, the actual call logs are used for analysis. These actual call logs are collected at MIT [6] by the Reality Mining Project group for a period of 8 months. The project collects mobile phone usage of 100 users which includes their user IDs (unique number representing a mobile phone user), time of calls, call direction (incoming and outgoing), incoming call description (missed, accepted), talk time, and tower IDs (location of phone users). These 100 phone users are students, professors, and staffs. We use this extensive dataset for our willingness estimator (WE) analysis and validation of 10 sample users in this paper. More information about the Reality Mining Project can be found in [6].

The main contribution of this paper is to propose the Willingness Estimator (WE) that can estimate the calles willingness without his/her involvement, the model estimates willingness level based on call history. In Section II we present the methodology and architecture of the Willingness Estimator that computes willingness level of the specified callee. Further, Section III describes the willingness computation. Finally, the accuracy of the proposed Willingness Estimator is validated with the actual call logs in Section IV.

II. Methodology

In our daily life, when we make a phone call, we often guess whether or not our call will be answered by the intended called party (callee). Most of the time, we want our call to be answered by the intended callee, however other times we want to leave a message instead of speaking to the callee. Therefore, when we make a phone call, we try to estimate our chance of being answered by the callee. We base this estimation on

Time of the day: The callee is not likely to take a call during his/her busy hours or while he/she is sleeping at night but more likely during his/her free hours such as time before work or during his/her break or driving home after work. Therefore, we estimate the callees willingness of taking a call based on when (time of call) we make the call.

Location: The callee is not likely to take any call while he/she is at work or in the theater but he/she more likely to take a call while he/she is at home or apartment. Therefore,

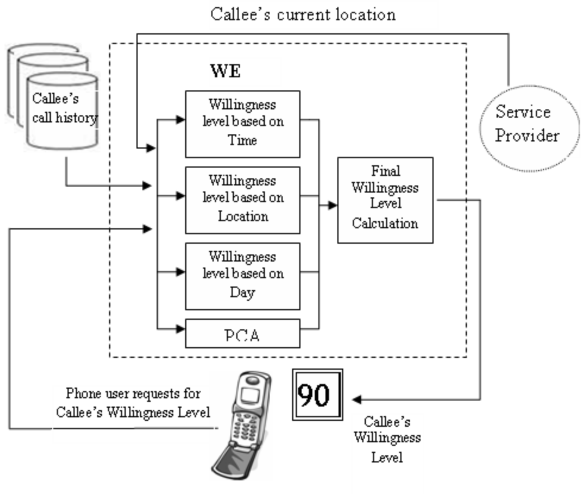

Fig. 1. Architecture of the Willingness Estimator.

we can also base our estimate of the callees willingness on where (location) the callee is located when we make a call.

Day of the week: Since we all have different schedules, most people go to work during weekdays and stay at home on the weekends. Thus, incoming calls during the weekends are more likely to be answered than during the weekdays. Likewise, we base our estimation of the callees willingness on what day of the week we make the call.

As previously mentioned, the caller wants to know the callees willingness of taking the call before the caller decides whether or not to make that call. *From the communication network controllers point of view, this can help traffic congestion since the caller knows the callees willingness level so the caller might not initiate a call which can reduce the traffic in the network and also save the callers available minutes.* Therefore, we propose the Willingness Estimator (WE) for computing the willingness level of the callee, which can be deployed at the base station. The basic architecture of the WE and its service flow diagram are shown in Fig. 1 and Fig. 2 respectively.

When the phone user makes a request to the WE for the callees willingness level, the WE takes information of time of the call (current time), day of the call (current day) and the callees location information from the service provider and callees call history from database. The WE computes the willingness level based on time of the call, location of the callee, and day of the call. The principal component analysis (PCA) is applied to compute the amount of contribution (importance) of those three input parameters have toward the final willingness level. The final willingness level is computed and forwarded to the phone user. It is then up to the phone user to decide whether or not to initiate the call.

III. Willingness Computation

Willingness can be defined as a wanted activity. This refers to receiving wanted calls in case of voice call. As described in Sec. 2, willingness level of the callee depends on time of the call, location of the callee, and day of the call.

We define the willingness level based on time of the call as the sum of the call frequency at that particular time divided by the total number of calls, which is given in Eq. (1) where $W_T(t)$ is the willingness based on time of call of t^{th} hour, N is the total number of calls, $n_t(i)$ is the j^{th} call frequency at t^{th} hour, and j = 1, 2, 3, , m where m is the total number of days of observation.

$$W_T(t) = \frac{1}{N}\sum_{j=1}^{m} n(i) \tag{1}$$

Similarly, the willingness level based on callees location is defined as the sum of the number of calls (call frequency) that the callee has received at a particular location divided by the total number of calls, which is given in Eq. (2) where $W_L(l)$ is the willingness based on callees location of l^{th} location, N is the total number of calls, and n_l is the number of calls at l^{th} location.

$$W_L(l) = \frac{n_l}{N} \tag{2}$$

Likewise, the willingness level based on day of the week $(W_D(d))$ is defined as the sum of the number of calls that the callee has received each day of the week (nd) divided by the total number of calls (N), which is given by Eq. (3). The sample plots of $W_T, W_L,$ and W_D are shown in Fig. 3(a), 3(b), and 3(c) respectively

$$W_D(d) = \frac{n_d}{N} \tag{3}$$

Therefore, the final willingness level (W) is given by Eq. (4).

$$W = C_T W'_T(t) + C_L W'_L(l) + C_D W'_D(d) \tag{4}$$

The final willingness level is the sum of the product of the normalized (rescaled to the maximum value) willingness level (W') based on time of call, callees location, and day of the call and the contribution coefficients $C_T, C_L,$ and C_D corresponding to time of call, callees location, and day of the week respectively. These contribution coefficients are introduced here because we believe that these three parameters (time, location, and day) have different contributions that impact the final willingness level. To obtain the values of these coefficients, PCA is applied.

Typically, PCA [7] is used to reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as mush as possible of the variation present in the data set. This is achieved by transforming to a new set of variables, the principal components (PCs), which are uncorrelated, and ordered so that the first few retain most of the variation present in all of the original variables. Scree plot was developed by Cattell [9] for selecting the number of PCs to be retained in order to account for most of the variation in X. The PCs are successively chosen to have the largest possible variance [7]. Suppose the variance of the k^{th}PC is l_k, scree test involves looking at a plot of l_k against k and

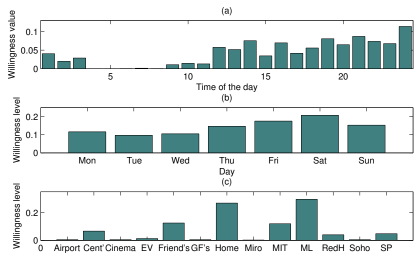

Fig. 2. (a) A sample willingness level based on time of the call. (b) A sample willingness level based on day of the week. (c) A sample willingness level based on location.

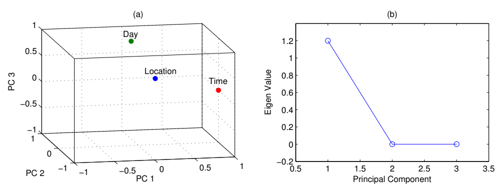

Fig. 3. (a) Principal component plot where Time of the call, Location of the callee, and Day of the call are represented with red, blue, and green dots respectively. (b) Scree plot shows that the first principal component has the highest eigen value.

TABLE I
PRINCIPLE COMPONENT MATRIX SHOWS THE VALUES OF TIME OF THE CALL, CALLEES LOCATION, AND DAY OF THE CALL ON THE PRINCIPAL COMPONENTS.

Parameter	**PC1**	PC2	PC3
Time of the call	**1.00**	0.00	0.00
Callee's location	**0.40**	0.92	0.00
Day of the call	**-0.08**	0.04	0.99

deciding at which value of k the slope of line joining the plotted points are steep to the left of k, and not steep to the right. This value of k defining an elbow in the graph is then taken to be the number of PCs to be retained. We apply PCA analysis to find the contribution of each parameter towards the final willingness level based on time of the call, callees location, and day of the call. We convert the three dimensional dataset into three PCs, and further look into the first PC which captures maximum variance of the data. Figure 4(a) shows where the three parameters (time, location, and day) lay on the principal component plot.

The first PC captures maximum variance as it has the highest Eigen value and the other two components can be eliminated from the analysis as referred to the Scree plot in Fig. 4(b); so area of interest lies in the first PC. Table 1 shows that on the first PC, the orders of significance from high to low is time of the call, callees location, and day of the call respectively. The physical significance of PCA is to find the underlying pattern in the dataset, and detect each fields contribution. Hence the numerical values of the contribution coefficient can be computed from Table I as the ratio of values the time, location, and day lied on the first principal component. For this sample the particular user has contribution coefficients of time, location, and day as 67.613%,

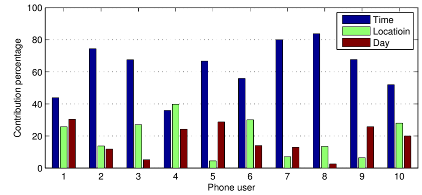

Fig. 4. Contribution coefficient plot of 10 different phone users resulting of the PCA

27.04%, and 5.206% respectively. Thus time factor plays a major role for this user as the majority of the calls are received during a particular time period as it can be seen from Fig. 3(a). So the contribution of time towards the willingness value for this user is high, whereas the location has the second highest contribution while day of the week has the lowest significance. Based on our analysis, we discover that PC values vary from person to person as PC captures individuals behavior.

IV. VALIDATION

To evaluate the accuracy of the model proposed in Section III, an experiment is conducted with actual call logs of 100 phone users (described in Sec. 1) over a period of 8 months. We randomly select 10 phone users who are students, professors, and staffs to represent all users. The first 6 months of data are used to compute the initial willingness level. The following 2 months are then used to validate the proposed Willingness Estimator (WE) where the willingness level is recomputed for each incoming call and verified against the result of the call (missed or accepted).

The accuracy is measured by the unwanted rate over the range of different willingness levels. The unwanted rate is the ratio of the number of missed calls to the total number of calls at the given willingness level. An assumption is made here that a missed call is considered as an unwanted call.

Fig. 4 shows the resulting contribution coefficients by applying the PCA for 10 different users to compute the contribution that each field makes towards the final willingness level. We observe that the contribution coefficients are different for all users. It can be observed that the phone users 1, 4, 6, and 10 (who are students) have tendency of taking calls at any time of the day thus time contribution coefficients towards willingness are relatively less than those users 5, 7 and 9 (professors) and users 2, 3, and 8 (staffs).

The numerical results are shown in Table II and the graphical representation is illustrated in Fig. 5. As we can see in Table II, unwantedness is higher in the low willingness regions (less than 50%) than the high willingness (more than 50%) regions. Again, the proposed WE only provides the willingness level of the specified callee; however it is up to the phone user to make a decision whether to initiate call. Nevertheless, the experimental results show high accuracy of our WE. Further, results validate our hypothesis that when the estimated willingness level is high, most of the calls are answered, whereas few calls are answered at the low willingness level.

TABLE II
EXPERIMENTAL RESULTS

Phone User User	No. of Incoming calls	No. of Unwanted calls	Unwanted rate(%) Willingness levels(%)									
			0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
1	541	117	15.38	48.71	19.65	15.38	6.83	3.45	0.00	0.00	0.00	0.00
2	438	106	24.52	20.18	13.20	9.43	14.00	10.00	6.60	2.07	0.00	0.00
3	210	28	28.57	0.00	14.28	28.57	14.30	14.28	0.00	0.00	0.00	0.00
4	620	176	14.00	29.00	11.50	6.25	10.50	15.00	10.00	3.75	0.00	0.00
5	134	46	26.08	32.78	32.78	0.00	0.00	0.00	8.36	0.00	0.00	0.00
6	88	30	20.00	60.00	13.30	0.00	6.70	0.00	0.00	0.00	0.00	0.00
7	104	59	5.09	57.63	18.64	0.00	18.64	0.00	0.00	0.00	0.00	0.00
8	861	128	12.50	14.06	55.47	17.97	0.00	0.00	0.00	0.00	0.00	0.00
9	170	28	7.21	28.50	64.29	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10	169	17	80.00	10.00	10	0.00	0.00	0.00	0.00	0.00	0.00	0.00

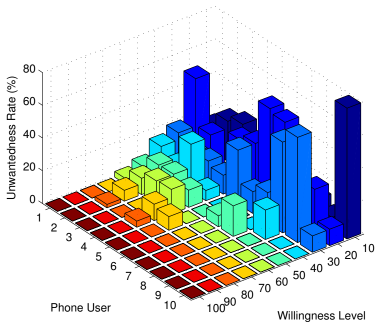

Fig. 5. Experimental results show the unwanted rates over the range of willingness levels for 10 different phone users.

V. CONCLUSION

In this paper, we proposed a Willingness Estimator (WE) that computes the willingness level of the callee. We propose a special button on the phone called presence. Before making a call, one can press this button for finding out if the calleee is indeed available. Next, it is up to the phone user to make a decision whether to initiate a call. The willingness computations based on these three factors are carried out as a simple ratio of the call frequency at particular (time, location, day) to the total number of calls. We believe that contribution of these factors to the overall (final) willingness level are different. It is validated by the principal component analysis that these three factors have different relevant weights to the final willingness level. Hence, the final willingness level of the callee is the sum of the product of the willingness based on time of the call, callees location, and day of the call, and its corresponding contribution coefficients. The accuracy of the model is evaluated with the actual call logs of 10 phone users from the MITs Reality Mining data sets. The WE performs well with high accuracy as when computed willingness level is low; the unwanted rate is high and vice versa. We believe presence will be a new service offered by service providers. If the willingness level is high, it is proved that the call will be answered but when the willingness level is low, the call is most likely to reach the callees voice mail. Most of the time, we want to speak to the callee but some other times we wan to leave a more meaningful message therefore this service will give the phone user the ability to predict the reaction of the callee after the phone ring. This service can also be useful for sales people to know when to make an important call to the customers. This service can also help the call traffic congestion by reducing unwanted communication traffic. In reality, there are many other factors that affect the willingness of the callee such as mood (state of mind), social relationships (between callee and caller), situation or status of the callee (e.g., out of available minutes, out of battery, callee has a second line, callee is in emergency, etc.) To quantify these factors is a very challenging task. We are working on other estimation techniques for improving the accuracy of the WE. These issues, among others, will be be addressed by our future work.

ACKNOWLEDGMENT

This work is supported by the National Science Foundation under grants CNS- 0627754, CNS-0516807, CNS-0619871, and CNS-0551694. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. We would like to also thank Nathan Eagle at MIT Media Lab for providing us the datasets.

REFERENCES

[1] J. Rosenberg, "Programming IP Telephony Services with the Call Processing Language (CPL) and CGI," *OPENSIG*, 1999.

[2] D.M. Smith, M.C. Cain, L. Latham, B. Burton, "Develop a Strategy for Presence Technology," *Gartner Research*, 2006.

[3] D. Jiang, T.H. Yeap, L. Logrippo, R. Liscano, "Personaliztion for SIP Multimedia Communications with Presence," *IEEE ICSSSM*, 2005.

[4] V. Gehlot, A. Hayrapetyan, "A Formalize and Validated Executable Model of the SIP-Based Presence Protocol for Mobile Applications," *ACMSE*, Winston-Salem North Carolina USA, 2007.

[5] X. Shan, A. Shriram, "Enterprise Mobile Applications Based on Presence and Logical Proximity," *IWCMC*, Vancouver British Columbia Canada, 2006.

[6] Massachusetts Institute of Technology: Reality Mining. http://reality.media.mit.edu/.

[7] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. Springer Science+Business Media, 1986, New York USA.

[8] I.T, Jolliffe, B. Jones, T. Morgan, "Cluster analysis of the elderly at home: a case study," *Data analysis and Informatics*, pp.745757, 1980.

[9] R.B.Cattell, and S.Vogelmann, "A Comprehensive trial of the scree and KG criteria for determining the number of factors," *Mult. Behav. Res.*, vol.12, pp.289-325, 1977.