Inferring User Relationship from Hidden Information in WLANs

Ningning Cheng, Prasant Mohapatra, Mathieu Cunche, Mohamed Ali Kaafar, Roksana Boreli, V. Srikanth

▶ **To cite this version:**

Ningning Cheng, Prasant Mohapatra, Mathieu Cunche, Mohamed Ali Kaafar, Roksana Boreli, et al.. Inferring User Relationship from Hidden Information in WLANs. MILCOM - IEEE Military Communications Conference - 2012, Oct 2012, Orlando, United States. hal-00747850

HAL Id: hal-00747850

https://inria.hal.science/hal-00747850v1

Submitted on 2 Nov 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Inferring User Relationship from Hidden Information in WLANs

Ningning Cheng[1], Prasant Mohapatra[1], Mathieu Cunche[2],
Mohamed Ali Kaafar[2], Roksana Boreli[3], Srikanth Krishnamurthy[4]
[1]Department of Computer Science, University of California, Davis, [2]INRIA Rhône-Alpes Grenoble France,
[3]National ICT Australia, [4]Department of Computer Science and Engineering, University of California, Riverside
Email: {nincheng, prasant}@cs.ucdavis.edu mathieu.cunche@inria.fr
kaafar@inria.fr roksana.boreli@nicta.com.au krish@cs.ucr.edu

***Abstract*—With ever increasing usage of handheld devices and vast deployment of wireless networks, we observe that it is possible to collect data from mobile devices and reveal personal relationships of their owners. In the paper, we exploit the hidden information collected from WLAN devices and infer individual relationships between device pairs based on three observation dimensions: network association history, physical proximity and spatio-temporal behavior. By measuring WLAN data, we demonstrate that device owners with social relationship tend to share access points, or show similar behavior patterns in wireless communications (e.g. go to the same place periodically to access the same WLAN network). These results can be exploited for various network analytic purposes.**

I. Introduction

With the fast development of mobile computing and wireless communication, people are becoming more and more attached to the mobile handheld devices, such as smartphones, PDAs and tablets. The physical devices have been so tightly coupled to the network users, that the network structure is largely dependent on the distribution of network users and their relationships in the network. In order to provide better network designs and enhance network performance in the future, it is imperative to study the user behaviors and their relationships within the network. What makes relationship discovery more important is the fact that it provides critical information for network applications in some networks. Consider two networks as examples: the delay tolerant network (DTN) [1] and the tactical network [2]. In DTNs, each user works as a network router and disseminate the information in a store-and-forward manner. In this case, relationship discovery helps decide which two (or more) users will meet more often and hence optimize the routing strategy and expedite the information propagation. In tactical networks, members from different teams (such as spies or malicious members) can form an orchestrated group and communicate with each other periodically, resulting in disclosing classified information to the espionage organization. In this case, relationship discovery helps to identify hidden relations between the agents and hence reveal the covert communities inside the tactical networks.

In this paper, we focus on relationship discovery in WLANs. The study of relationship in WLANs faces challenges due to its unique characteristics. First, recent years have witnessed a significant growth of mobile WLAN users. It is easy for them to join or leave, which makes it difficult to infer relationships by simply taking network snapshots. Second, Mobile users can roam between different WLANs at different time under different names. The difficulty of tracing a mobile user brings challenge in the relationship discovery. Third, most private WLANs adopt encryption mechanism such as WEP or WPA to preserve data confidentiality. When users access the network through encrypted channels it is difficult to get relationship information by tapping the wireless media. These challenges motivate us to explore hidden information that is not generated from wireless communication channels, but from users' implicit behavior patterns. It is supported by the fact that society is formed by the congregation of people with similar behaviors. Therefore, in mobile wireless networks, people with similar mobility patterns should have a stronger social tie.

In order to discover user relationships, we focus on the similarity of users' behavior patterns. Our first observation is that similarity of SSIDs of previously accessed networks implies user relationships. Since most of the private WLAN network are encrypted, users who are able to access the same private network share the same key, hence are very likely to know each other. Users that access same public networks in the past may also have relationship if they share multiple common networks. Therefore, the similarity of devices' network access history can be used to infer relationship between the device owners. Our second observation is that users who locate in the same building and access the network from the same location are more likely to be related to each other, or have potential relationship. For example, it is common for one organization to have more than one department and each department having its own network for the employees. Even though the employees from different departments have different network to access, they may still know each other from work collaborations. Hence we make physical proximity as the second observation dimension. Third, we assume that users with high temporal similarity are more likely to have relationships. For example, friends and family members meet more often than strangers. In this case, we observe the spatio-temporal co-occurrence frequency of the users and generate *spatio-temporal* similarity to infer their relationships.

The rest of this paper is organized as follows. Section 2 defines the problem and our proposed framework. Section 3 presents experimental set up and some measurement based refinements. Section 4 demonstrates the results of our experiment, Section 5 discusses related work and finally, Section 6 concludes the paper.

II. Relationship inferring framework

In this section, we define the problem, set up the notations and definitions, and introduce the framework of our solution.

A. Problem statement

The goal of this paper is to leverage the information that can be observed from portable wireless devices (e.g. notebook, netbook, tablet, pda and smartphone) and infer possible relationship between the device users.

Social groups by nature consists of individuals with similar behavioral patterns. Based on this observation, in order to infer user relationships, we explore WLAN users' behavioral similarity from three aspects: the similarity of previously accessed networks; the proximity of user locations and the frequency of co-occurrence.

Since the meaning of relationships can be multi-faceted and context-dependent, we clarify that our work mainly focus on relationships that are related in real life. For example, users who are friends in online social networks but not know each other in the real life are not considered in this paper. Our work does not intend to discover user relationship with certainty. Its goal is to narrow down users that may have relationships from a large sample poll, or strengthen conjectures such as the existence of a relationship between some users. Instead of self-reporting data, our method is observation based. Therefore it can detect objective relationships such as working interactions or neighborhood relationship (where communication can be observed).

B. Exploration user behavior and similarity metrics

1) Network association similarity: Current technology has made it possible to get the network access history of Wi-Fi devices. To speed up the process of reconnecting to the WLAN access points, most operations systems (e.g. Windows, Mac OS, Linux, iOS and Android) keeps a Preferred Network List (PNL) of previously accessed network names. When a wireless device is discovering the WLAN network, the default setting is to first actively probe for the previous accessed network names by their Service Set Identifier (SSID). The PNL also decides the order of the SSIDs being probed. The SSID information is contained in the Probing Request Frame (PRF), which is broadcast in plain text before any encryption mechanism is applied. A wireless network adapter will keep requesting for the SSIDs based on certain order until some AP replies with a probe response frame. This SSID list is very user specific and be used to uniquely identify a user [3].

In this paper, our first step is to compare user similarity based on their network access history. Since a device always broadcast SSIDs in plain text, the information is public accessible to anyone. Only after this phase, device starts to exchange authentication packets with AP, and encrypted the communication channel.

Given two devices d_1 and d_2 and their previous network access lists n_1 and n_2, their similarity can be compared [4] based on certain similarity metric (such as Jaccard, Pearson or Cosine similarity metric). In this paper, we will use Cosine metric because it is claim to outperform other existing similarity metrics [4], [5].

$$Similar(d_1, d_2)_a = \frac{\overline{n}_1 \cdot \overline{n}_2}{\|\overline{n}_1\| \, \|\overline{n}_2\|} \tag{1}$$

where $\overline{n}_1$ and $\overline{n}_2$ are n_1 and n_2 normalized by the corpus of all SSIDs.

2) Location proximity: According to the first law of geography, (everything is related to everything else, but near things are more related than distant things [6]), the mobile users' interactions between places are inversely proportional to the travel distance between them. Hence, geographic location proximity can be considered to further improve the relationship inference.

For location proximity consideration, we focus on the location of the Access Points (AP) mobile user has connected to. Since The Wi-Fi devices are portable and some are highly attached to the user, the SSID list that shows previously accessed networks also implies that user has been to the places where the networks locate. In this case, locations of the device's probing SSIDs reflect a large sample of the user's location history. We use network-location coupling to further identify the similarity between mobile users. As long as the AP's location is given, the user's location history is revealed. With online AP database such as WiGLE [7], the AP's name can be mapped into geographical coordinates that reveals the location history of the user.

In order to check if network-location mapping is needed, we will check the collocation of two devices before applying the SSID similarity metric.

$$Need\_Mapping(d_1, d_2)_l = \Sigma\eta(n_i, n_j) \tag{2}$$

where n_1 and n_2 are the SSID lists for d_1 and d_2, $n_i \in n_1$, $n_j \in n_2$ and $\eta(n_i, n_j)$ checks if n_1 and n_2 have collocated SSIDs. $\eta(n_i, n_j) = 1$ if n_i, n_j in same location; $\eta(n_i, n_j) = 0$ if n_i, n_j in different locations. Any non-zero result means the network-location mapping is needed.

3) Spatio-temporal co-occurrence probability: Mobile users may demonstrate periodic reappearances at certain locations. Users who are related are more likely to gather together or meet frequently than unrelated strangers. Thus we make spatio-temporal co-occurences the third aspect of inferring social relationships.

Spatio-temporal co-occurence is defined as the probability that the user u_1 and u_2 occur together at the same place and time. Each user's behavior can be modeled as a temporally distributed process at different places, with random variables representing the user's reoccurrence frequency at that location during different timeslots. Hence, we assign each device a matrix D, with the column representing the location and rows representing the timeslots we capture this user. For example,

$$D = \begin{pmatrix} 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 2 & 1 & 0 & 0 \\ 0 & 1 & 1 & 2 \end{pmatrix} \tag{3}$$

An entry $D(i, j)$ represents the number of reoccurrence that device show up at corresponding times in the corresponding location. Then the temporal similarity which describes the co-occurrence of a pair of devices is defined as follows.

$$Similar(d_1, d_2)_o = \sum_j \frac{D_{1j} \cdot D_{2j}}{\|D_{1j}\| \, \|D_{2j}\|} \tag{4}$$

where the temporal similarities at different locations are summed up to compare occurrence similarity.

TABLE I
EXAMPLE OF A WLAN USER PROFILE

MAC address	SSID	Location	Timeslot
a1:b2:c3:d4:e5:f6	attwifi	starbucks	1pm-2pm
a1:b2:c3:d4:e5:f6	hello	starbucks	1pm-2pm
a1:b2:c3:d4:e5:f6	lisa's network	Bldg1	3pm-4pm

TABLE II
AN EXAMPLE OF INFERRING RELATIONSHIP FROM THREE SIMILARITY METRICS

Relationship	SSID similarity	SSID sim with loc consideration	Spacial temporal similarity
no	1.6E-6 (weak)	1.6E-6 (weak)	0.1 (weak)
yes	3.4E-3 (strong)	1.2E-2(strong)	0.45 (strong)
yes	0.1234 (strong)	1 (strong)	0.6 (strong)

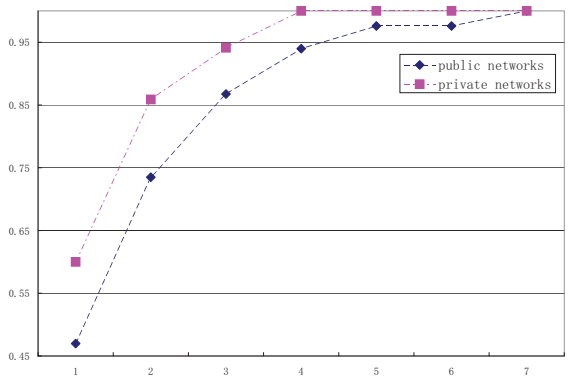

Fig. 1. CDF of ssid frequencies in public and private networks in semi-log scale

III. EXPERIMENTAL SET UP AND REFINEMENTS

A. Data capture

In the experiment, we set our device's Wi-Fi interface to monitor mode and passively monitor the WLAN probing request frames within our communication range. The experiment is done at four campus hotspot locations during four rush hours for one month. We record the time-stamp, the source MAC address, the location and SSIDs being probed and use them as the Wi-Fi device's profile. Table I shows an example of a user's profile with hypothetical information. Then we examine the similarity of user profiles on three aspects and infer social relationships based on the combined knowledge of similarities (one example is shown in Table II).

In the experiment, we observe several facts that can lead to bias or inaccurate inference of similarities due to the characteristic of Wi-fi probings. The SSID list device probes records the previous networks the device has accessed to. However, there are several problems we need to address.

Our first observation is that two pair of nodes with same number of common SSIDs can shows different tie strengths. For example, if the SSID is a public network name commonly used in different places(e.g. *"attwifi"* is used for most starbucks APs), the users' relationship can be weaker. If the SSID only belongs to a home network which is unique in the world, the users are supposed to have stronger relationship. In order to differentiate kind of networks and give high importance to unique network name, it is necessary to assign different weights to different SSIDs.

Another observation is that different network service platforms provide different strategy of sending probing request packets. The request can be sent in order of recently accessed order or longest connection time, or only request for networks that are accessed in the last month. As soon as the Wi-Fi device receives the probe response frame from the AP, it stops broadcasting probe frames and starts to communicate with the AP. Therefore, the SSID list we can capture is highly dependent on how new the environment to the users. A new user will give more SSID information than an old (a regular) user. As far as we know, Windows system sends probing information in the order of most recently accessed networks. As long as it receives the probe response frame from the AP, it stops broadcasting probe request frames and start to associate with the AP. This observation leads to the result that the SSID list we collect can be partial information. One method to overcome this problem is capturing the SSID from different environments. In our experiment, we focus our data capture in four different hot spots. The SSIDs collected from the same device in different places during different timeslots will be merged if the lists are different.

B. Association history similarity

In reality, SSID has different meanings and lead to different strength of relationship. There are SSID names like *"linksys"* and *"comcast"*, which are the default SSIDs given by the router's manufacturer $Cisco$ or the Internet service provider $Comcast$. There are campus or enterprize SSIDs like *"UC-Davis"* and *"eduroams"*, that are shared by multiple APs in the same institution or company. There is also unique SSID name that are used by certain user in private networks (like *"lisa's network"*). The tie strength of relationship differs based on which kind of network the users are sharing. In order to give high weights to specific and unique SSID, we assign a weight for each SSID.

SSID weight assignment: We examine the frequency (f) of different type of network names and discover one of the main difference is their frequency of being probed. For example, as shown in Figure 1, public networks are being probed more frequently than private networks. Therefore, in order to show the importance of different SSIDs, we assign each SSID a weight which is inverse proportional to its frequency of being probed.

Then according to Equation 1, we calculate two devices' SSID similarity as follows:

$$Similarity(d_1, d_2)_a = \frac{\Sigma \beta_z^2}{\sqrt{\Sigma \beta_i^2} \sqrt{\Sigma \beta_j^2}} \tag{5}$$

$$\beta = \frac{1}{f} \tag{6}$$

where d_1, d_2 refer to $device_1$ and $device_2$, β is the weight of an SSID, which is inverse proportional to its frequency f, z is the set of common SSIDs both in d_1 and d_2s' preferred network list, i and j are the SSID lists of d_1 and d_2 respectively.

In order to find the similarity threshold for SSID metric, we trained a control set that maximize the True Positive Rate (TPR) and minimize the False Positive Rate (FPR). Here

True Positive (TP) (resp. False Positive (FP)) is the number of related pairs (resp. unrelated pairs) that are inferred to have relationship in our method. Similarity, True Negative (TN) (resp. False Negative (FN)) is the number of unrelated pairs (resp. related pairs) that are not inferred to have relationship in our method. TPR is defined as $TP/(TP+FN)$, reflecting the sensitivity of our method. And FPR is defined as $FP/(FP+TN)$, reflecting the (1- specifity) of our method. The controlling set is based on 20 volunteers' relationships. Of all the 190 links between 20 volunteers, we use 90 of them as the training set and the other 100 relationships as the testing set. Figure 2 shows the Receiver Operating Characteristic (ROC) curve of detected relationship. According to the result, we choose our threshold $3.05E-4$, where TPR is 0.76, FPR is 0.18 and TPR/FPR is maximized.

Based on this threshold, we can calculate each pair of devices' SSID similarity and discover potential relationship between device owners.

C. Similarity with location considered

For location measurement, we detect the existing networks in each campus building and group the AP names in the same building as one cluster. For example, if $Bldg_1$ has two SSIDs $SSID_1$ and $SSID_2$, we will map them into same location $Bldg_1$. For future representation, devices looking for either $SSID_1$ or $SSID_2$ are considered to have been in the same location $Bldg_1$. In this case, by comparing the number of buildings where the devices accessed the network, we get location similarity of two users.

$$Similarity(d_1, d_2)_l = Similarity(M(n_1), M(n_2))_a \quad (7)$$

where M is the function that maps SSIDs into their geographic locations, n_1 and n_2 is the SSID list of d_1 and d_2.

Location proximity serve as complementary information for SSID-based relationship detection. Consider co-workers at same layer of building who know each other. If they access the network from their own labs by different APs, they will not have common SSIDs. Hence SSID-based metric will lose this relation. On the other hand, location-based similarity will merge their lab SSIDs into single building and hence discover the relationship between them.

With activities such as Wardriving (persons mapping wifi networks by a mobile vehicle, using a portable computer or smartphone), it is possible to get public AP maps from some wireless network database. In this paper, we use the AP map from an online database [7], to group the SSIDs we collected from on campus access points and group them into 25 campus buildings. Figure 3 shows a snapshot of the AP maps of University of California, Davis from [7]. The red dots in the map represent the APs and the SSID names of the APs are also given in the database. Based on this, we can set up a mapping table from SSID names to the building names.

D. Spatio-temporal similarity

Co-occurrence is another aspect for the study of user relationships. In reality, whether two person meet often is an indispensable information to infer if they are related to each other. People performing social behaviors like meetings or group discussion requires encounter with each other. The repetition and duration of encounters reflect how strong the users' relationship is.

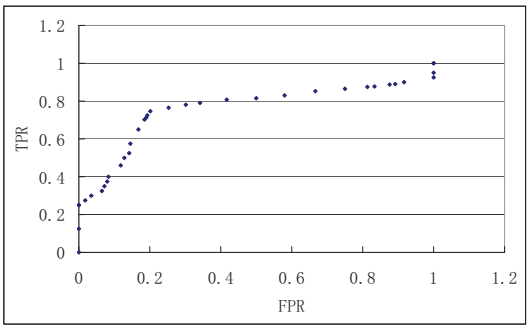

Fig. 2. The ROC curve of TPR and FPR of detected relationship in the training set by SSID similarity

We use spatio-temporal similarity to describe users co-occurrence behaviors. We collect users' wireless activity by monitoring if their device generate packets during certain timeslots. Then we can infer user relationship by exploiting the devices' co-location history and encounter history. Without a complete deployment of monitoring system, we can only get partial information from the network. However, as long as we get enough sampling of users' trace data, we can discover potential relationships from these partial information. The estimation is based on the similarity of user profiles in WLANs we capture. A fake mobile users spatio-temporal profile is shown in Table I. Each entry of the trace has the location of association and session time duration information for that user.

In our experiment, we passively pick up packets at four locations that users most frequently go to (one starbucks, one cafeteria and two student activity centers). The timeslot is set to one hour. We record the probing history at four rush hours (12pm-2pm, 4pm-6pm) for one month. And put the number of time we observed a user show up into a 4 by 4 matrix. This matrix represent this user's spatio-temporal profile. We use similarity metric introduced in Equation 4 to calculate two users' spatio-temporal similarity, where the column is the location, the rows is timeslots and each entry is the users show up frequency.

IV. Results: relationship inferences

A. Relationship inference based on SSID similarity

We detect possible relationship in 30 days from the wild dataset we collected. The result is validated by randomly picking 10-12 users (including 45-66 relationships) from the test set and mix them with the wild dataset, as shown in Table III. Note that our method largely reduces the sampling pool and detect users pairs that share at least one common preferred networks. And the FPR is controlled under 25% based on the threshold we trained in the previous section. However, false negative results cannot be eliminated because some public area can have more than one access points with different names. Therefore, even two users access the same network with different AP names, they are inferred to be unrelated. In this case, we need other metric to detect user relationships from a different aspect.

B. Relationship inference based on location similarity

To further improve the relationship inference, we use the location similarity to explore more possible relationships. We

TABLE III
NUMBER OF DETECTED RELATIONSHIPS AND VALIDATION OF THE RESULT

Device number	Detected relationship	testing set	TPR	FPR
959	2008	66	0.734	0.166
623	1596	45	0.896	0.0892
853	987	55	0.798	0.176
934	1659	45	0.725	0.182
983	2015	55	0.733	0.206
739	883	66	0.815	0.15
839	1012	55	0.798	0.176
881	832	45	0.725	0.182
932	2304	66	0.734	0.166
739	934	55	0.733	0.206
853	987	45	0.815	0.125
602	539	66	0.815	0.15
746	1248	55	0.798	0.176
821	1012	45	0.733	0.186
638	828	45	0.810	0.106
849	1430	45	0.715	0.166
651	843	55	0.833	0.162
792	1504	45	0.725	0.182
834	1721	55	0.744	0.202
684	923	55	0.798	0.176
710	878	55	0.818	0.145
539	832	55	0.733	0.206
758	1280	55	0.798	0.176
592	660	45	0.803	0.194
749	892	55	0.733	0.130
605	773	66	0.825	0.180
638	801	55	0.833	0.162
733	1120	55	0.798	0.176
838	1837	45	0.725	0.182
664	822	55	0.744	0.202

couple public SSID names with the campus buildings it belong to, so whenever two users probe for two different SSIDs from the same building, they are mapped to the same location hence can generate a stronger relationship than other users. For this experiment, we collect all the SSIDs that show up in 25 main campus buildings (representing 38 different departments), as shown in Figure 4, and map different SSIDs within same building into one location.

Before calculating the SSID similarity between a pair of devices, we first check if the their SSIDs can be mapped to the same department building. The SSIDs should only belong to this building. If the mapping is successful, we assign a potential relationship to the pair of devices. Otherwise, only the SSID similarity is measured. Result shows that with location information, we can detect 30% more relationships than simply using SSID metric (Figure. 5).

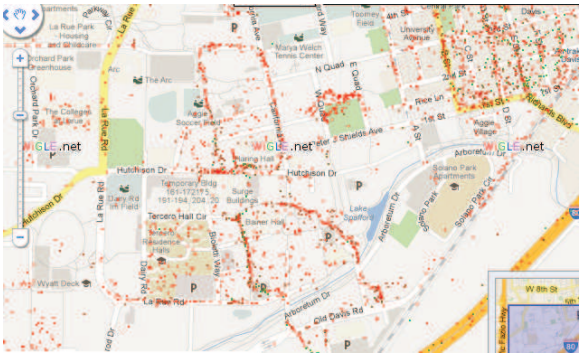

Fig. 3. Map of APs in UC Davis, each dot represents a Wi-Fi access point on the campus

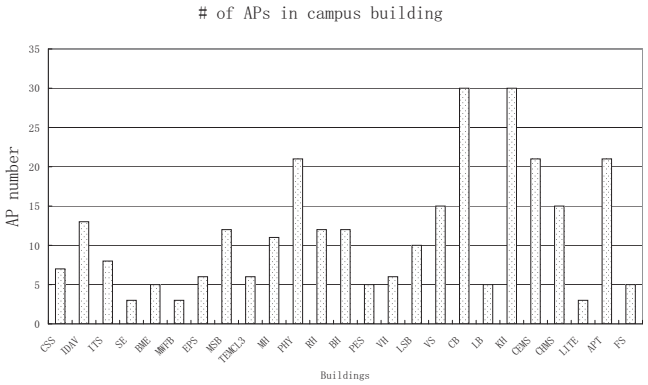

Fig. 4. Number of different APs in same campus buildings

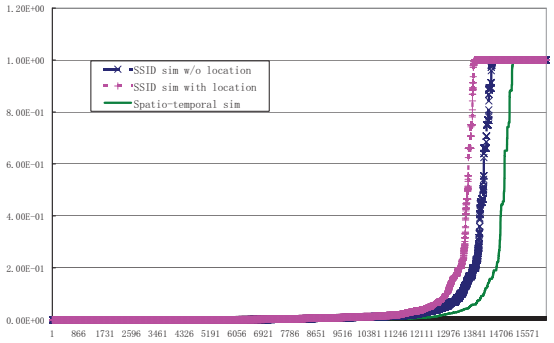

Fig. 5. SSID similarity, location similarity and spatio-temporal similarity

C. Relationship inference based on spatio-temporal similarity

The results for spatio-temporal similarity are shown in Figure 5, where detected related pairs of users are plotted in the ascending order regarding their similarity value. From this result, we find potential pairs of related users by examining the spatio and temporal overlaps. It shows spatio-temporal is another complementary dimension for relationship discovery.

In order to validate the result of spatio-temporal similarity, we pick all the devices that have linkages based on spatio-temporal similarity metric, and calculate their SSID similarity both with and without location consideration (Figure 6's (a) and (b)) in four locations. Result shows that over 70% of the devices can also be linked by SSID similarity. For the devices that can be detected based on both SSID similarity and spatio-temporal similarity, we examine their their average and variance of the SSID similarity over 30 days and discover the similarity result is quite stable (Figure 6 (c), (d) refer to results without location information, and (e), (f) with location information). Hence validates spatio-temporal similarity which is coherent with SSID similarity.

V. RELATED WORK

Relationship inference in online social network has been well discussed in recent years. Based on information content, such as emails or blogs, relationship can be drawn from communication archives or message traffics [8]–[12]. One of the earlier approaches in relationship discovery is to set up a generative model to discover correlation or dependency between entities. In this case, the relationship is substantiated by the content of the information data and the information traffic

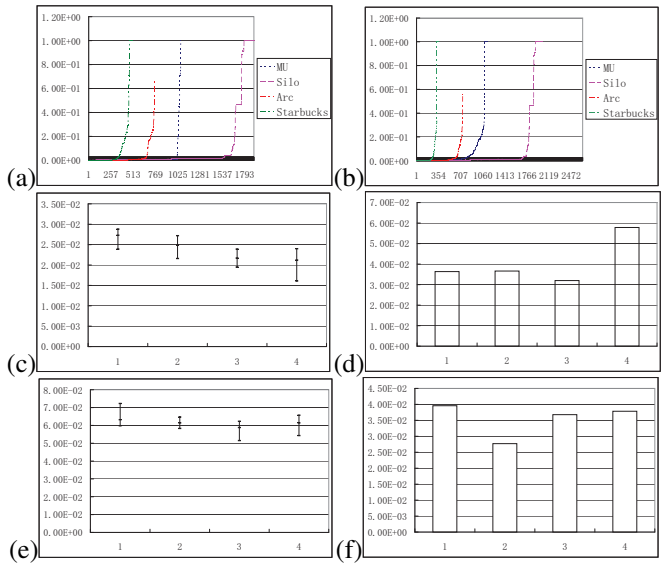

Fig. 6. For devices detected with spatio-temporal metric, their distribution of the SSID similarities w/o and with location consideration is shown in (a) and (b). Their average and variance of SSID similarity w/o and with location consideration is shown in (c, d) and (e, f).

between the users. Another approach is to infer relationship from the network structure [13], [14]. Different from previous approach, this one needs the complete network structure.

Relationship discovery in mobile networks (e.g. WLAN, cellular network) has recently drawn researchers attention. Relationships such as user-user encounter or user-base-station encounter is largely dependant on the users' social behavior and can impact network performance by affecting network workload. In this case, instead of looking at users' communication content, the pattern of user behavior can be exploited to infer a social relationship. Cranshaw et al [15] studied the user behavior in WLAN traces and inferred objective relationship based on user profile similarity. Relationship inference based on behavior similarity is discussed as a new research area. Relationship discovery based on WLAN users' association logs is discussed in [16], [17]. In [18], a study of mobile phone data proves that similar behavior pattern in cell phone data can provide inference of user relationship. In this paper, Eagle et. al show that the observational cell phone data can generate friendship structures, which is in consistence with users' self-reported friendship structure.

VI. Conclusion

In this paper we have presented and analyzed user behavior in WLAN networks based on a trace collected at campus hotspots. The goal of our study is to extend the understanding of wireless users' relationship by comparing their behavioral patterns obtained from hidden information in WLAN networks. After characterizing wireless users in terms of network association history, geographic location proximity and spatio-temporal co-occurrence frequency, we compare the similarity of user behaviors in these three aspects and infer possible relationship from their respective similarity measurements. Our work can be applied to social community detections or social tie inference to understand WLAN users' grouping behavior. It can also improve the wireless network deployment and potential network optimizations in user-centric applications.

VII. Acknowledgements

Research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-09-2-0053. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

References

[1] P. Hui, J. Crowcroft, and E. Yoneki, "Bubble rap: social-based forwarding in delay tolerant networks," in *Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing*, MobiHoc '08, pp. 241–250, ACM, 2008.

[2] C. Weinstein, W. Campbell, B. Delaney, and G. O'Leary, "Modeling and detection techniques for counter-terror social network analysis and intent recognition," *2009 IEEE Aerospace conference*, pp. 1–16, 2009.

[3] J. Pang, B. Greenstein, R. Gummadi, S. Seshan, and D. Wetherall, "802.11 user fingerprinting," in *Proceedings of the 13th annual ACM international conference on Mobile computing and networking*, MobiCom '07, (New York, NY, USA), pp. 99–110, ACM, 2007.

[4] M. Cunche, M. A. Kaafar, and R. Boreli, "I know who you will meet this evening! linking wireless devices using wi-fi probe requests," in *13th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 2012.

[5] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. T. Riedl, "Application of dimensionality reduction in recommender system – a case study," in *ACM WEBKDD Workshop*, 2000.

[6] Q. Li, Y. Zheng, X. Xie, Y. Chen, W. Liu, and W.-Y. Ma, "Mining user similarity based on location history," in *Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems*, pp. 34:1–34:10, ACM, 2008.

[7] "http://www.wigle.net/,"

[8] H. Kautz, B. Selman, and M. Shah, "Referral web: combining social networks and collaborative filtering," in *Communications of the ACM*, vol. 40, pp. 63–65, ACM, Mar. 1997.

[9] M. Aida, K. Ishibashi, C. Takano, H. Miwa, K. Muranaka, and A. Miura, "Cluster structures in topology of large-scale social networks revealed by traffic data," *GLOBECOM 05 IEEE Global Telecommunications Conference 2005*, vol. 1, pp. 41–46, 2005.

[10] C. P. Diehl, G. Namata, and L. Getoor, "Relationship identification for social network discovery," in *Proceedings of the 22nd national conference on Artificial intelligence - Volume 1*, pp. 546–552, 2007.

[11] G. Kossinets, J. Kleinberg, and D. Watts, "The structure of information pathways in a social communication network," in *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '08, pp. 435–443, ACM, 2008.

[12] R. Xiang, J. Neville, and M. Rogati, "Modeling relationship strength in online social networks," in *Proceedings of the 19th international conference on World wide web*, pp. 981–990, ACM, 2010.

[13] J. Leskovec, K. J. Lang, and M. Mahoney, "Empirical comparison of algorithms for network community detection," in *Proceedings of the 19th international conference on World wide web*, WWW '10, pp. 631–640, ACM, 2010.

[14] M. E. J. Newman, "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 23, pp. 8577–8582, 2006.

[15] J. Cranshaw, E. Toch, J. Hong, A. Kittur, and N. Sadeh, "Bridging the gap between physical location and online social networks," in *Proceedings of the 12th ACM international conference on Ubiquitous computing*, Ubicomp '10, pp. 119–128, ACM, 2010.

[16] W.-j. Hsu, D. Dutta, and A. Helmy, "Mining behavioral groups in large wireless lans," in *Proceedings of the 13th annual ACM international conference on Mobile computing and networking*, MobiCom '07, pp. 338–341, ACM, 2007.

[17] G. S. Thakur, A. Helmy, and W.-J. Hsu, "Similarity analysis and modeling in mobile societies: the missing link," in *Proceedings of the 5th ACM workshop on Challenged networks*, CHANTS '10, pp. 13–20, ACM, 2010.

[18] N. Eagle, A. S. Pentland, and D. Lazer, "Inferring friendship network structure by using mobile phone data," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 36, pp. 15274–15278, 2009.