Interest-aware Implicit Multicast (iCast): Opportunistic Mobile Data Dissemination without Per-group Management

Saeed Moghaddam, Ahmed Helmy
Computer and Information Science and Engineering Department
University of Florida, Gainesville, FL
{saeed, helmy}@cise.ufl.edu

ABSTRACT
The unprecedented tight coupling between mobile devices and their users provides new opportunities to infer users' behavior and interest. Much of the future mobile services shall center on human behavior and interests, and will rely heavily on the paradigm of multicast communications, as in group communication in mobile social networks. In this paper, we introduce a new casting paradigm, interest-aware implicit multicast (iCast) that works based on the inferred interest profiles. In this paradigm, messages are sent to a behavioral interest profile (not to an IP or device address). In this manner it is considered opportunistic where the users are implicitly matched based on their profiles using co-clustering algorithm. There is no need for explicit member join or leave, and no per-group membership management. We explore a spectrum of architectures including centralized, semi- centralized and distributed and conduct a campus-wide case study for the evaluation of the second architecture. Within this architecture, we suggest two different message delivery approaches including semi and full interest-aware; and show how the later approach can significantly improve the performance utilizing a set of recommendations that are built based on a behavioral interest model of mobile society.

Categories and Subject Descriptors
C.2.1 **[Computer-Communication Networks]**: Network Architecture and Design—Wireless communication.

General Terms
Algorithms, Performance, Design, Experimentation,

Keywords
Interest profile; Behavioral targeting; Mobile recommendation.

1. INTRODUCTION
Wireless mobile networks are evolving and integrating with every aspect of our lives. Today, laptops, handhelds and smart phones are becoming ubiquitous providing almost continuous Internet access. This significant shift toward mobile Internet access has been accelerating with the rise of larger multi-touch smartphones and tablets by providing better and easier Internet access experiences than previous generations of mobile devices. In fact, the usage of mobile Internet is progressing so fast that it is going to revolutionize the entire framework of communication technology. In the last few years, not only has the use of cell phones increased in quite a dramatic way, but the way that people prefer to utilize them, communicate and stay in touch with the world has changed too. People today are ever-increasingly utilizing online services on the move using their mobile devices for different purposes, e.g., listening to music, watching videos, sending and receiving emails, web browsing, and social networking. This fast growing trend toward mobile Internet access creates a tight coupling between users and mobile networks where various characteristics of user online activities, mobility and interests can be captured and applied to provide novel solutions to the networking problems.

Mobile data dissemination is one of the major services, which can be re-designed based on the acquired user characteristic via the mobile network. In a mobile network where human behavior and interests can be implicitly inferred via their online activities and mobility, a new paradigm can be developed for opportunistic interest-aware casting of information. In such a paradigm messages are delivered based on users behavioral interests not their networking device addresses. However, current Internet abstractions (i.e., unicast, multicast) do not lend themselves readily to interest-aware mobile networking. The nature of such communication paradigms is based on network characteristics not humans. On the other hand, today's interest-based services rely on elaborate information and settings explicitly provided by users, hindering the effective participation of classes of users including seniors, children, and handicapped. Therefore, we need a paradigm shift in design of mobile networks to be able to provide effective and practical interest-aware services in a mobile social network. In such a new approach, implicit inference of users' interests shall be the first step toward the design of opportunistic interest-aware communications.

Mobile networks now have enough potential to accurately capture and infer multi-dimensional aspects of users' behavior, interests and preferences. Mobile devices can now act as distributed behavioral sensors of users to capture their interests and enable implicit interest profiling. Such enriched profiles can be then analyzed for various mobile networking scenarios at various spatio-temporal granularities to facilitate the creation of accurate behavioral interest models. Based on the implicitly acquired interest profiles and interest models, we can develop a new type of opportunistic communication paradigm that we refer to as *iCast*. ICast tries to deliver messages to multiple mobile users based on their implicitly inferred interests.

This work was partially funded by NSF awards 1320694, 0832043.

MobiArch'14, September 11, 2014, Maui, Hawaii, USA.

http://dx.doi.org/10.1145/2645892.2645894

In this paper, we first discuss different types of architectures which can be used to develop the iCast paradigm including centralized, semi-centralized, distributed and hybrid architectures. Then, we show how such a paradigm affects the performance of networking communications. For this purpose, we consider different approaches of information delivery in a mobile network including semi and full interest-aware approaches and analyze their performance. The rest of the paper is as follows. Section 2 reviews related work. Section 3 introduces iCast paradigm and architectures. Section 4 presents a case study analysis results for a semi-centralized architecture and Section 5 concludes the paper.

2. RELATED WORK

On analysis and modeling of mobile networks, there has been a widespread interest. The scope of analysis includes WLAN usage and its evolution across time [1-3] and user mobility [4, 5]. The work on the TVC model [6] provides a data-driven mobility model for protocol and service performance analysis. In [7] it was shown that the performance of resource scheduling and TCP vary widely between trace-driven analysis and non-trace-driven model analysis. Several other works focus on classifying users based on their mobility periodicity [8], time-location information [9], or a combination of mobility statistics [10] and using user behavior characteristics to design realistic and practical mobility models [11-13]. They have shown that most widely used existing mobility models (mostly random mobility models, e.g., random walk, random waypoint; see [14] for a survey) fail to generate realistic mobility characteristics observed from the traces which are essential for protocol performance [15].

The two main trace libraries for the networking communities can be found in the archives at [16] and [17]. None of the available traces provides large-scale netflow information coupled with DHCP and WLAN sessions to be able to map IP addresses to MAC addresses to AP to location and eventually to a context (e.g., history department). In this study, we use our dataset, including the above traces, which is significantly larger and richer in semantic than the other wireless mobile network traces and includes around 100 million records per day.

On mobile data dissemination and casting, there has been some earlier works. Profile-cast [7] provides a one-to-many communication paradigm targeted at a behavioral groups. In the profile-cast paradigm, messages are sent to those who match a behavioral profile. However, the profiles are only based on location visitation preferences and are not aware of web access patterns. Moreover, the casting paradigm in profile-cast is not based on multiple aspects of user interests. Other works also rely on movement patterns.

In the previous works, we proposed different techniques for multi-aspect modeling of users' interest based on co-clustering [18], self-organizing maps [19-21], Gaussian mixture model [22], domain/location specific [23] and global/local modeling [24]. In other works, frequent pattern mining is used to provide relevant information to the users on their mobile devices [25, 26]. In this work, we introduce a novel paradigm which uses enriched interest profiles based on web and location visitation patterns and utilizes advanced behavioral modeling techniques for targeting and disseminating the messages.

3. INTEREST-AWARE CASTING (iCAST)

3.1 Data-driven Design Paradigm

Realistic design of services and protocols for mobile networks requires a deep understanding of dynamics within mobile society. To obtain such an insight, we need to collect, process and model extensive multi-dimensional data regarding users online activities and mobility before starting the design process. We refer to this approach as data-driven design. The eventual goal of the data-driven design paradigm is to utilize analysis of users' behavior and interest to drive the design of efficient interest-aware protocols and services. This approach starts by realistic analysis and modeling of users' interests (based on their online activity, mobility, etc.) that then drives the design process.

Realistic analysis and modeling of large wireless mobile networks requires processing of multi-dimensional datasets with fine granularity. For this purpose, extensive datasets need to be collected, using network infrastructure or devices, that may be augmented by online directories (e.g., buildings maps) and web services (e.g., whois lookup service). This huge amount of data then needs to be processed in order to cross correlate the obtained information from different sources (e.g., IP and MAC addresses) and get the data integrated and aggregated in an appropriate format. The aggregate data is then used to build user interest profiles and to model the mobile society. The acquired interest profiles and models provide the building blocks for the interest-aware design process.

3.2 Interest-aware Networking

Today's Internet uses machine-oriented abstractions (e.g., IP addresses) for routing unicast and multicast traffic. Moreover, users are required to explicitly input their membership and interests information. Such architecture creates inefficiencies due to sub-optimal indirections, and complex group management. Therefore, new sets of mechanistic building blocks are needed to provide efficient interest-aware networking without the explicit collection of users' interests.

We propose a new network abstraction that embodies the notion of interest-based addressing. Instead of sending a message to a destination address (unicast or multicast IP) the message is destined to a target interest profile. This notion is general and a profile can be defined as needed. We provide concrete behavioral interest profiles based on mobility and online activities. Figure 1, shows a sample representation for the interest profile. In this case, the user interest profile shows the percentage of his online time at different buildings for different web domains over a specific period of time.

In order to develop the idea of interest-aware casting, different architectures can be considered including centralized, semi-centralized, and fully distributed architectures. A common set of essential building blocks is needed for all these architectures, including: i) interest-profile builder that consists of data collectors and profile processors, ii) interest-aware dissemination that consists of target identification, interest matching, and interest-based routing. In the following, we discuss these architectures and their corresponding building blocks.

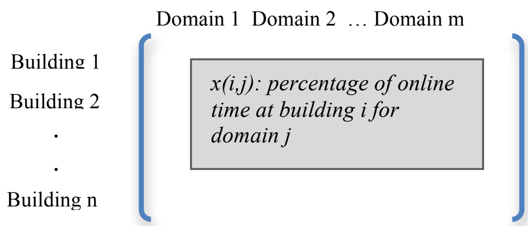

Figure 1. A sample representation for behavioral interest profile.

3.3 Centralized Architecture

This architecture (Figure 2) is used to illustrate the role of the behavioral interest profile abstraction. The behavioral data may either be collected by access points in the infrastructure, or reported by the users' mobile devices. The data is then processed (mined and stored) by a centralized server (or a group of servers) and behavioral interest profiles are generated. The process of sending a message starts by target identification (to be used as the destination profile) using behavioral interest metrics (e.g., target users with main interest in "netflix" and "mac" websites who visit the "library" and "business fraternity" frequently). A sender generates the target profile representation, and then sends the message and the target to the server for matching and dissemination. Realization of this scheme may borrow from existing centralized services.

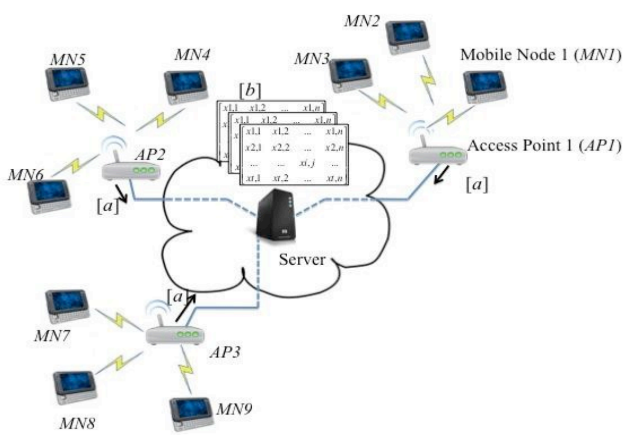

Figure 2. Centralized architecture: a. Wireless APs collect and relay to server data about activity/mobility of users. b. Server constructs activity/mobility interest matrices for all users and processes them for similarity matching. Issues of reporting overhead and privacy arise.

3.4 Semi-centralized Architecture

In this architecture (Figure 3), data processing is done partially at the users' mobile devices, and only a general interest profile representation (showing high level behavioral interest, but not details) is sent to the server to reduce privacy concerns. This profile representation not only preserves user privacy to some extend but also provides enough information for target matching process for the messages at the server side.

3.5 Distributed Architecture

In the distributed architecture (Figure 4), mobile devices not only generate the behavioral interest profiles, but also disseminate and route the messages in a peer-to-peer fashion by locally performing the target profile matching. As no data is stored and processed on a server, this scheme preserves privacy but may have lower performance. Therefore, a hybrid approach may also be considered depending on the required service and performance, and availability of infrastructure. For example, semi-centralized architecture may be used during normal operation while distributed scheme may be utilized during emergencies when servers fail.

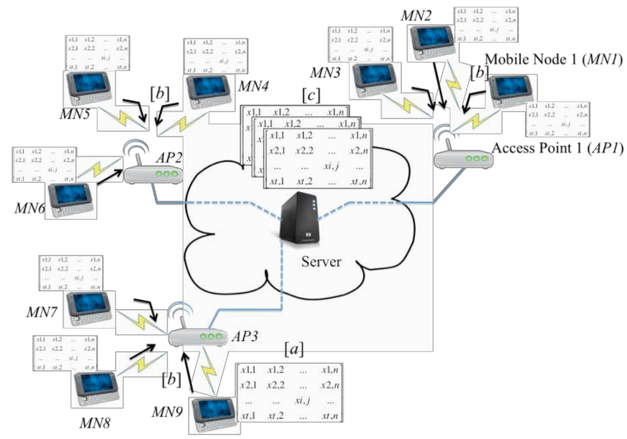

Figure 3. Semi-centralized architecture: a. Wireless devices (MNs) collect and store activity and mobility interest matrices for individual users. b. The information is relayed to the server occasionally. c. Server processes information for all users for similarity matching.

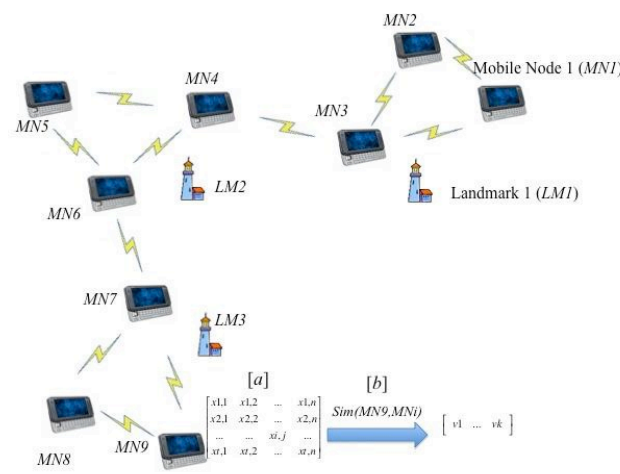

Figure 4. Distributed architecture: *a.* Wireless devices (MNs) collect and store activity and mobility interest matrices for individual users. *b.* Similarity processing is performed locally at each device during message forwarding. No reporting of profiles or servers is required.

4. CASE STUDY: SEMI-CENTRALIZED iCAST

In our case study, we simulated the semi-centralized architecture for interest-aware casting. For this purpose, we collected users' activity and mobility logs from a campus wireless network infrastructure and then got it processed to form users' interest profiles. We used these offline-generated interest profiles instead of locally generated profiles by mobile devices in the actual architecture. In the simulation, we assigned these profiles to the users which represent the real mobile society. Using this simulated environment, we ran the semi-centralize interest-aware casting as explained before. In the following, we explain each of these steps in detail.

4.1 Collection of Activity and Mobility Logs

We collected different types of extensive traces via network switches (in USC campus) including netflow, DHCP and wireless session logs. An IP flow is defined as a unidirectional sequence of packets with some common properties (e.g., source IP address) that pass through a network device (e.g., router) that can be used for flow collection. Network flows are highly granular; flow records include the start times and duration, source and destination IP addresses, port numbers, protocol numbers, and flow sizes. The IP addresses can be used to identify user device MAC addresses using DHCP log and the websites accessed respectively. The DHCP log contains the dynamic IP assignments to MAC addresses. The wireless session log collected by each wireless AP includes the 'start' and 'end' events for device associations (when they visited or left that specific AP) and can be used to derive the location of users at any time.

4.2 Profiling of Users Interests

The variety and scale of different collected traces introduces one of the main challenges with respect to data processing. The netflow dataset gathered from the campus includes around 2 billion flow records for each month which equals to 2.5 terabytes of data per year. To circumvent the problem, we first compressed and exported the data into a database management system and then designed customized stored procedures for data integration (mapping source IPs to MAC addresses (user IDs) and destination IPs to domain names). Then, in order to build the users' interest profiles, we aggregated the integrated data based on user ID, domain name and building and calculate the total online time for each resulting record. In our case study, we created users interest profiles for all mobile users across campus (22,816 users) considering all their interactions during a month with top 100 active web domains over 79 different buildings.

4.3 Interest-aware Message Delivery

According to the semi-centralized architecture, when user profiles are built on the mobile devices, these profiles are sent to a central server. On the other hand, when a user wants to send a message, he needs to send the message as well as its target profile to the server. When server receives the message, it needs to determine target users for the message to be sent. For this purpose, we suggest two interest-aware approaches that we refer to as *semi interest- aware* and *full interest-aware* casting. These approaches use user interest profiles to find the matched users for the massages. In the following, we explain these two approaches and compare their performance with a none-interest-aware approach, which we refer to as *random cast*.

4.3.1 Random Cast

Flooding is the simplest way to do the message dissemination. If we do not collect and form user interest profiles, the only way to transmit the messages to the interested users is to send every single message to all users and let them decide if they want to read that message or not. It is obvious that using this approach, puts a huge burden on the mobile network, mobile devices and of course users to transmit, receive and select the desirable/useful messages. To remedy this issue, an immediate approach is to send messages to only a random portion of users not all depending on the available resources and the number of messages. We refer to the later approach as random cast.

4.3.2 Semi Interest-aware Casting

In the semi interest-aware approach, target users are selected based on user interest profiles that are maintained by the server. For each message, the server examines all the available profiles and sends the message to only top matched users (considering a threshold). This approach as we will discuss later, can improve the performance of message delivery that we measure based on similarity of receivers' profiles with the target profile determined by the sender.

4.3.3 Full Interest-aware Casting

In this approach, we not only utilize users' interest profiles on the server, but also try to build a big picture of the mobile society based on available profiles and provide this big picture in the form of a set of recommendations to the senders. The rationale behind this idea is the fact that when senders have no idea about the community interests, they might frequently try to target users with specific interests who actually do not exist in the society. Providing a big picture of the community interests helps them to target the receivers. For this purpose, senders are provided with a set of recommendations describing different existing interest groups inside the mobile society. In this way, senders can get a good insight on people interests inside the mobile society and choose an appropriate target for their messages.

However, to take this approach we first need to find the behavioral interest groups inside the mobile society. For this purpose, we need to create realistic models of users' interests using their interest profiles. However, large scale of users and high dimensionality of their interests (they might look at a vast verity of websites for example) are a major challenge to achieve this goal using basic clustering approaches, e.g., hierarchical clustering or k-means. In our previous studies [18], we have shown that information theoretic co- clustering model [27] can be effectively applied to solve this problem in large mobile societies. This model can provide a multi-dimensional characterization of dynamics within mobile society considering web and location visitations. In this model, users as well as their profile features (i.e., their interests on websites or locations) are clustered simultaneously. Hence, we can identify different behavioral groups representing similar points of interest (considering websites, locations, etc.). This model has been shown to be stable around 90% over several months obviating the need to rerun the method frequently. In the following, we briefly explain this modeling approach.

4.4 Co-clustering Interest Model

The input for the co-clustering algorithm, are users interest profiles, which represents the amount of their online time at different websites or locations. Information-theoretic co-clustering technique [27] treats the non-negative contingency table as a joint-probability distribution of two discrete random variables, whose values are given in the rows and columns, and poses the co-clustering problem as an optimization problem in information theory. This technique defines mappings from rows to row-clusters and from columns to column-clusters and then tries to optimize the co-clustering result. The optimal co-clustering is one that leads to maximum mutual information between the clustered random variables, and minimizes the loss in mutual information between the original random variables and the mutual information between the clustered random variables. This algorithm monotonically increases the preserved mutual information and optimizes the loss function. This task is performed by intertwining both row and column clustering. Column clustering is performed by calculating closeness of each column distribution (in relative entropy) to column cluster prototypes. Row clustering is performed similarly. This iterative process converges to a local minimum. The algorithm never increases the loss, and so, the quality of co-clustering improves gradually. Iteratively, the method performs an adaptive dimensionality reduction and estimates fewer parameters than one-dimensional clustering approaches, resulting in a regularized clustering. In addition, the algorithm is efficient. The computational complexity of the algorithm is given by $O(N \cdot \tau \cdot (k + l))$ where k and l are the desired number of row and column clusters, N is the number of non-zeros in the input joint distribution and τ is the iterations.

4.5 Evaluation and Analysis

For the evaluation purpose, we measure similarity of receivers' interests and the targeted user interest profiles by the sender. This metrics shows how well the process of message delivery is performed. If the amount of similarity is high it shows that receivers will be actually interested in the received message but if not it shows the opposite. To measure this metrics we simulated the three mentioned approaches using our dataset. In the experiments, we assumed that a sender wants to send a message to a target profile. In the first two approaches, i.e., random and semi interest-aware, target might be any arbitrary profile, but in the third approach, i.e., full interest-aware, it can be one of the recommended ones. For each case, we repeated the experiment 10 times to remove any noises and measured the average performance of message delivery based on the similarity metric.

Figure 5 (random cast) shows the empirical CDF for interest similarity between receivers profile and an arbitrary target profile in random cast. X-axis shows similarity and y-axis shows the CDF value (representing the percentage of users with a similarity score less than or equal to the value on the X-axis at each point). In the experiment, a message with an arbitrary target was sent to a group of randomly chosen users (10 percent of population in this case). The experiment was repeated 10 times for different arbitrary targets and the plot shows the average result. As can be seen in the figure, in this approach more than 85 percent of receivers have a similarity score less than 40 percent. If we look at the maximum and average similarity score we also find the maximum similarity score that we can achieve in this approach is around 66 percent while the average is less than 27 percent. This shows that the random cast approach shows a poor performance in terms of similarity metrics.

Figure 5 (semi-iCast) shows the result for the semi interest-aware casting. The plot in the figure shows the empirical CDF of interest similarity between receivers profile and an arbitrary target profile. Again, this value is the average for 10 experiments. In each experiment, we matched the target against all the available profiles and send the message to only top 10 percent of users showing the highest similarity score. As can be observed form the plot as well, in this approach the maximum similarity score is the same as the random cast (66 percent), which is trivial, however the minimum similarity is increased to 38 percent. Therefore, the average similarity score for the receivers in this approach is improved from 27 percent in random cast to more than 45 percent. While this approach shows relatively good improvement, it still suffers from two issues. First, the amount of similarity score is not very high. Second, it does not improve the upper bound for the similarity score (we get the same 66 percent). This is very important because it shows that even if we decrease the threshed from top 10 percent to top 0.01 percent and increase the number of messages by factor of 1000, which will put a heavy burden on the network, we will not gain a similarity performance better than 66 percent.

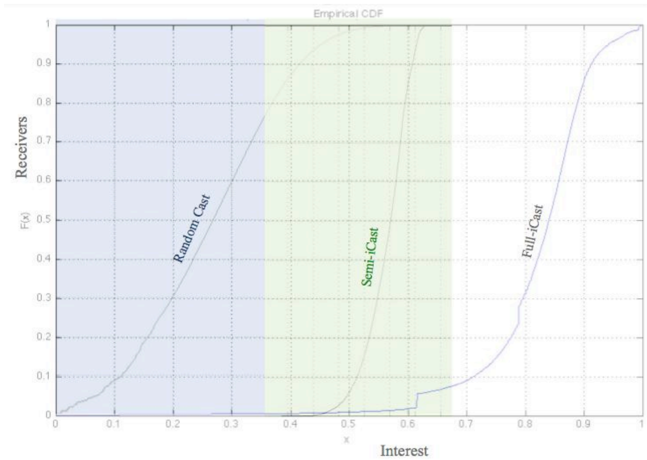

Figure 5. Empirical CDF for interest similarity between receivers profile for random cast, semi-iCast and full-iCast. X-axis shows interest similarity and y-axis shows the CDF value.

Figure 5 (full-iCast) shows the result for the full interest-aware casting. In the experiment, we had 10 recommended interest groups that were built using co-clustering method. The plot in the figure shows the average empirical CDF for interest similarity between receivers' profile and a recommended target profile. As can be observed form the figure, this approach does not improve the minimum similarity, however only less than 2 percent of receivers have a similarity score less than 60 percent. In this approach the maximum similarity score is close to 1 and the average is more than 82 percent. This shows that the full interest-aware approach can significantly improve the performance comparing to the other two approaches.

The later approach shows better result considering two other factors of scalability and usability as well. While in semi interest-aware casing, for each new message every single profile need to be examined to find the best match, which will put a huge computational burden on the server specially if we need to deal with millions of users an millions of messages, in the full interest-aware approach, we just need to build the interests groups once a while (with a specific frequency e.g. a day or week) and then reuse those groups to create recommended target profiles for the sender. Therefore, the latest approach shows a significant improvement in terms scalability as well. On the other hand, providing recommendations for the senders makes it a lot easier for them to choose the target groups comparing to the semi interest-aware approach. In the full interest- aware approach, senders can easily pick one of the recommended options even from their mobile devices without spending too much time, while in the semi interest-aware approach they need to set many detailed parameters for any single elements of target profiles (i.e. the percentage of online time in 100 different websites in this case) which is of course not very easy and pleasant.

5. CONCLUSION

This study is motivated by the need for developing new information casting paradigm for current and future mobile networks. For this purpose, we introduced a novel interest-aware implicit multicast approach for opportunistic mobile data dissemination called iCast. In the new paradigm, messages are sent to targets based on users' interest not their device or internet addresses. To develop such a paradigm, we proposed different architectures including centralized, semi-centralized and hybrid out of which we simulated the second one. For the simulation of semi-centralized architecture we proposed three different approaches for targeting the users including random, semi interest-aware and full interest-aware approaches. We evaluated the performance of these approaches based on a similarity metric showing the level of users interest in the received messages. We showed that the full interest-aware approach can significantly improve the performance of user targeting from 27 percent for random cast and 45 percent for semi-interest-aware approaches to 82 percent. We also discussed how the later approach can significantly improve the scalability and usability in a real application. In future, we are planning to implement the full interest-aware approach for the mobile devices and evaluate the performance of the proposed approach in a real world application.

6. REFERENCES

[1] Tang D. and Baker. M., Analysis of a local-area wireless network. In ACM MobiCom, 2000.

[2] Kotz, D., Essien, K. Analysis of a campus-wide wireless network. Wirel. Netw., 11, 1-2 (2005).

[3] Henderson, T., Kotz, D. and Abyzov, I. The changing usage of a mature campus-wide wireless network. Computer Networks, 52, 14 (Oct 2008), 2690-2712.

[4] Balazinska, M. and Castro, P. Characterizing mobility and network usage in a corporate wireless local-area network. ACM MobiSys 2003.

[5] McNett, M. and Voelker, G. M. Access and mobility of wireless PDA users. SIGMOBILE Mob. Comput. Commun. Rev., 9, 2 (Apr 2005), 40-55.

[6] Hsu, W., et., A. TVC: Modeling spatial and temporal dependencies of user mobility in wireless mobile networks. IEEE/ACM Trans. Netw., 17, 5 (2009).

[7] Hsu, W., Dutta, D. and Helmy, A. CSI: A Paradigm for behavior-oriented profile-cast services in mobile networks, Ad Hoc Networks 10(8): 1586-1602 (2012).

[8] Kim, M. and Kotz, D. Periodic properties of user mobility and access-point popularity. Personal Ubiquitous Comput., 11, 6 (Aug 2007), 465-479.

[9] Eagle, N. and Pentland, A. Reality mining: sensing complex social systems. Personal and Ubiquitous Computing, 10, 4 (May 2006), 268.

[10] Tang, D. and Baker, M. Analysis of a metropolitanarea wireless network.Wirel.Netw.8,2/3 (2002), 107-120.

[11] Jain, R., Lelescu, D. and Balakrishnan, M. Model T: a model for user registration patterns based on campus WLAN data. Wirel. Netw., 13, 6 (Dec 2007), 711-735.

[12] Lelescu, D., Kozat, U. C., Jain, R. and Balakrishnan, M. Model T++: an empirical joint space-time registration model. ACM MOBIHOC 2006.

[13] Kim, M., Kotz, D. and Kim, S. Extracting a mobility model from real user traces, IEEE INFOCOM 2006.

[14] Bai, F. and Helmy, A. A Survey of mobility modeling and analysis in wireless adhoc networks, Wireless Ad Hoc and Sensor Networks, Springer, 2006.

[15] Bai, F., Sadagopan, N. and Helmy, A. The IMPORTANT framework for analyzing the Impact of Mobility on Performance Of RouTing protocols for Adhoc NeTworks. *Ad Hoc Networks*, 1, 4 (2003).

[16] Kotz, D. and Henderson, T. Crawdad: A community resource for archiving wireless data at dartmouth. IEEE Pervasive Computing (Dec 2005), 12-14.

[17] MobiLib: http://nile.cise.ufl.edu/MobiLib/.

[18] S. Moghaddam, A. Helmy, S. Ranka, M. Somaiya, Data-driven co-clustering model of Internet usage in large mobile societies. ACM MSWiM 2010: 248-256.

[19] Moghaddam, S. and Helmy, A., Multidimensional modeling and analysis of wireless users online activity and mobility: A neural-networks map approach. ACM MSWiM 2011: 401-408.

[20] Moghaddam, S. and Helmy, A., Internet usage modeling of large wireless networks using self-organizing maps. IEEE MASS 2010: 612-617.

[21] Moghaddam, S. and Helmy, A., Spatio-temporal modeling of wireless users internet access patterns using self-organizing maps. IEEE INFOCOM 2011: 496-500.

[22] Moghaddam, S. and Helmy, A., SPIRIT: A simulation paradigm for realistic design of mature mobile societies. IEEE IWCMC 2011: 232-237.

[23] Moghaddam, S., Carvalho, M., Helmy, A., Domain and location specific modeling of mobile users online Interests. IEEE IWCMC 2013: 1436-1441.

[24] Almutairi, A., Somaiya, M., Moghaddam, S., Ranka, S., and Helmy, A., A global local modeling of internet usage in large mobile societies. ACM MSWiM PM2HW2N '12, 69-76.

[25] Srinivasan, V., Moghaddam, S., Mukherji, A., Rachuri, K., Xu, C., and Tapia, E.M. On-device mining of mobile users' co-occurrence patterns. HotMobile 2014.

[26] Srinivasan, V., Moghaddam, S., Mukherji, A., Rachuri, K., Xu, C, Tapia, E., MobileMiner: Mining your frequent behavior patterns on your phone. ACM UbiComp 2014.

[27] Dhillon, I. S. and Guan, Y. Information theoretic clustering of sparse co-occurrence data. IEEE ICDM 2003.