Autonomous and Distributed Recruitment and Data Collection Framework for Opportunistic Sensing

Güliz Seray Tuncay Giacomo Benincasa Ahmed Helmy
Department of Computer and Information Science and Engineering
University of Florida
Gainesville, FL 32611
{gstuncay, giacomo, helmy}@cise.ufl.edu

ABSTRACT

People-centric sensing is a novel apporach that exploits the sensing capabilities offered by smartphones and the mobility of users to sense large scale areas without requiring the deployment of sensors in-situ. Given the ubiquitous nature of smartphones, people-centric sensing is a viable and efficient solution for crowdsourcing data. In this work, we propose a fully distributed, opportunistic sensing framework that involves two main components which both work in an ad hoc fashion: Recruitment and Data Collection. We analyzed the feasibility of our distributed approach for both components through preliminary simulations. The results show that our recruitment method is able to select 66% of the nodes that are appropriate for the sensing activity and 88% of the messages sent by these selected nodes reach the sink by using our data collection method.

Categories and Subject Descriptors

C.2.4 [**Computer-Communication Networks**]: Distributed Systems—*distributed applications*; C.2.2 [**Network Protocols**]: Routing Protocols

General Terms

Design, Experimentation, Algorithms

Keywords

Crowdsourcing, DTN, Mobile Computing, Opportunistic Sensing, People-centric sensing

1. INTRODUCTION

Recent advances in sensing capabilities, storage capacity, and computation power of smartphones–along with the increasing pervasiveness of their adoption–led to the development of people-centric applications that exploit the capabilities of these devices to crowdsource data collection about the users' surroundings. People-centric sensing has therefore become a viable approach for large-scale urban sensing activities that do not require the deployment of sensors in-situ. Lane et al. discern two classes of people-centric sensing activities: participatory and opportunistic sensing [1]. Participatory sensing requires the participants to *consciously opt to meet the application requests* [1]; therefore, it is necessary that the participants are engaged in the sensing activity, or that the participants are rewarded to achieve better involvement. Conversely, in opportunistic sensing, minimal or no interaction with the participants is required; in fact, opportunistic sensing applications may run in the background and opportunistically collect data with no need to interact with the participant. High participation is crucial for the success of people-centric activities. Opportunistic sensing may be a suitable and inexpensive way to achieve high participation; however, current approaches for people centric-sensing rely on centralized registries to recruit possible participants [2]. Moreover, the collected data is uploaded from the participants to remote servers by using cellular networks. These approaches may lead to privacy and economical concerns that may work as a disincentive for participation. This paper describes a fully distributed and privacy-preserving framework for opportunistic sensing. The framework includes a recruitment component and a data collection component, both of which work in an ad hoc fashion and do not rely on the presence of an infrastructure. The only exceptions are the nodes that are designated as data sinks for collecting the sensed data, and the nodes designated as recruiters for the data collection campaign. The number of these nodes is expected to be orders of magnitude less than the number of sensing nodes. The framework consists of two main components: *recruitment* and *data collection*, both of which rely on previous work on profile-cast [3].

MobiCom'12, August 22–26, 2012, Istanbul, Turkey.
ACM 978-1-4503-1159-5/12/08.

2. PROFILE-CAST OVERVIEW

Profile-cast [3], is a behavior-oriented protocol that aims to send messages to the nodes matching a certain target *behavioral profile*, rather than explicit IDs, in opportunistic networks. Each node can autonomously construct its profile by monitoring the places it visits and the time spent at those locations. Based on this information, the node can build an association matrix that represents the percentage of time the node spent at various locations. Singular value decomposition (SVD) is then applied to this association matrix to obtain the dominant set of eigen-behavior vectors that represent the behavioral profile of the user. A target profile is a summarized mobility preference vector which is represented using the same format as that of the user behavioral profile. The similarity of two behavioral profiles is determined by computing the cosine similarity between their eigen-behavior vectors. In profile-cast, information is routed in two phases: *gradient ascend* and *group spread*. During the gradient ascend phase, messages are diffused following an ascending gradient of similarity between the behavioral profile of the node and the target profile. No replication is allowed; the message is transfered from a node to another with higher similarity to the target profile. When the message reaches a node with a higher similarity to the target profile than a certain threshold th_{sim}, the protocol switches to group spread mode. Group spread mode allows replication, and messages are copied from the holder to an encountered node with a higher similarity than th_{sim}. This double-phase of operation allows profile-cast to achieve a fair trade-off between overhead and delivery ratio.

3. THE OVERALL FRAMEWORK

3.1 Scenario

Profile-cast can be used for both recruitment and data collection. In the context of recruitment, specifically-designated nodes can be deployed prior to the start of a sensing activity and can disseminate recruitment messages that advertise the upcoming sensing activity. Since only the nodes that are likely to appear in the sensing area during the sensing time need to be recruited, the target profile of the recruitment messages will contain the locations that are included in the sensing area. For data collection, the sensing nodes need to send the collected data to the deployed data sinks. In this case, the messages can be sent using a target profile that describes the behavioral profile of the sinks.

3.2 Recruitment

The aim of the recruitment phase is to notify the nodes of an upcoming sensing activity. We exploit the stability of the user behaviors and based on the mobility history information, we recruit only the nodes that are likely to be in the sensing area when the sensing activity is taking place. In order to efficiently achieve a distributed recruitment, we propose the deployment of one or more recruiting nodes that visit the sensing area prior to the start of the sensing activity and then disseminate recruitment messages. Profile-cast can be used to perform the dissemination, and a profile which includes the locations in the sensing area can be used as the target profile. However, unlike profile-cast, where the aim is to reach nodes matching a certain target profile, for recruitment it is enough to reach the nodes that match only a *part* of the target profile. In other words, if the sensing area is large and consists of multiple locations, nodes that visit only few of the sensing locations should also be recruited. It is therefore necessary to change the similarity metric to be able to evaluate the similarity between single locations and the behavioral profile of the nodes. Finally, since the recruiting nodes are deployed in the sensing area, they can be configured to operate only in group spread mode; thus, they hold a copy of the message after replicating it to other nodes.

3.3 Data Collection

Once data is collected, it needs to be returned to the organizer of the sensing activity. We propose the use of data sinks for the collection of user data and suggest that the participating nodes opportunistically exploit ad hoc encounters to reach data sinks that are temporarily deployed in the sensed area. The data sinks are assumed to be temporary and mobile; moreover, they are not expected to be available for the whole duration of the sensing activity. In many cases, it is plausible that the locations visited by the data sinks are known or explicitly selected prior to the start of the sensing activity. This information can be included in the recruitment message and therefore can be used as the target profile to reach the sinks. While in profile-cast, the aim is to reach each nodes matching the target profile; in the context of data collection, the aim is to reach *any* of the nodes that match the target profile. For this reason, along with the fact that the sensing nodes are expected to greatly outnumber the number of data sinks and that they may produce a large number of messages, it is critical to limit the data replication by having the nodes run solely in gradient ascend mode. While this may significantly increase the delivery time, in many sensing applications it is not an issue.

4. EVALUATION AND RESULTS

In order to evaluate our approach, we developed a trace-driven simulation using USC Wi-Fi traces, in a way analogous to what is described in [3]: users are assumed to be in communication range when they are associated with the same access point at the same time. For the opportunistic sensing activity, we designed a test scenario in which the sensing activity is scheduled on a part of the USC campus. Since opportunistic sensing works best when there is high participation in the sensing activity, the sensing activity was scheduled to take place in the area including the seven most-visited locations. A single node, designated both as a recruiter and a data sink, is deployed one week before the start of the sensing activity, and continuously moves along a seven waypoint predefined path, that is assumed to be included in the recruitment message. The sink spends five minutes at each waypoint, before moving to the next. This behavior is continuously repeated along the duration of the sensing (from 8 am to 8 pm).

4.1 Recruitment

In our recruitment scenario, the recruiter node is configured to send a recruitment message every hour until the start of the sensing activity. The similarity threshold, th_{sim}, used in the recruitment simulation, is 0.8. A node is recruited if it receives the recruitment message before the sensing activity starts and is similar to the target profile in the recruitment message by at least 80%. We analyzed the traces and learned that the number of nodes that appear in the sensing area during the sensing activity, and hence should be recruited, is 376. The number of nodes which received the sink's recruitment messages is shown in Figure 1. 247 nodes are recruited for the sensing activity, which corresponds to 66% of the nodes that are eligible for taking part in the opportunistic sensing activity. It is worth highlighting that all 247 recruited nodes were among the 376 nodes that were eligible to take part in the sensing activity. Figure 1 also shows the overhead in terms of the number of nodes that are reached by the recruitment message. The reason for this inefficiency is that the behavioral profile is based on historical information and it is therefore possible that the behavior of the nodes is not always consistent with their mobility.

Figure 2 shows the compound number of recruited nodes as a function of the recruitment start time. It can be observed that the number of recruited nodes converges very quickly.

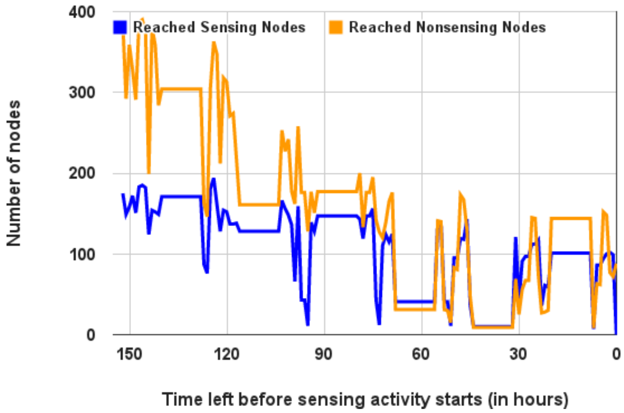

Figure 1: Number of nodes that received the recruitment message (recruited and not recruited) and number of nodes recruited (progressively) versus time

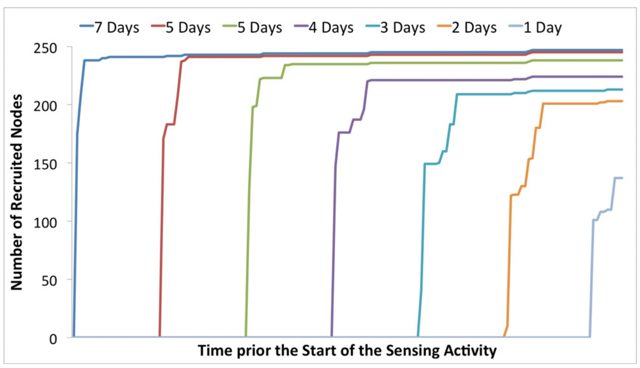

Figure 2: Number of nodes recruited (progressively) versus time

4.2 Data Collection

The nodes send their collected data every hour when they are in the sensing area. In the first scenario, we assumed ideal recruitment, meaning that all the nodes that are in the sensing area are recruited and take part in the sensing activity. Each of these nodes send one message to the sink per day. 2419 out of 2879 messages sent during the sensing activity are able to reach the sink node; that is, the delivery ratio of our method is 84%. In the second scenario, only the nodes that are selected by the recruitment algorithm send messages to the sink node. 1718 out of 1952 messages reached the sink, meaning that the delivery ratio of the proposed method is 88%. Figure 3 shows the messages sent by the nodes in the areas that are visited by the sink and the messages that are delivered. For each location, the first bar shows the number of delivered and undelivered messages sent from this location for the ideal recruitment case, while the second bar represents the case in which only the nodes selected by the recruitment phase send messages to the sink. It can be observed that a significant portion of the sent messages are delivered for both cases and the results for the second case are slightly better.

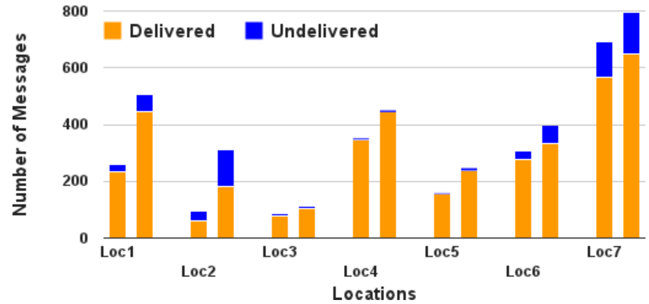

Figure 3: Delivery status of the messages from the locations visited by sink for 1)ideal recruitment case and for 2)only selected nodes.

5. CONCLUSION AND FUTURE WORK

In this work, we proposed a framework for autonomous and distributed recruitment and data collection in opportunistic sensing. For participant recruitment, we plan to investigate the addition of the progress review phase [2] into the framework and introduce online recruitment for selecting additional participants while the sensing is already taking place. For data collection, we plan to analyze the efficiency of appropriate unicast routing protocols and compare them with our proposed method. Additionally, we plan to consider recruiting sinks from the pool of users based on their mobility histories, by taking into account metrics such as behavior stability and variety of encounters. Furthermore, we plan to develop an opportunistic sensing application for smartphones in the future.

6. REFERENCES

[1] N. D. Lane, S. B. Eisenman, M. Musolesi, E. Miluzzo, and A. T. Campbell. Urban sensing systems: opportunistic or participatory? In *Proceedings of the 9th workshop on Mobile computing systems and applications*, HotMobile '08, NY, USA, 2008. ACM.

[2] S. Reddy, D. Estrin, and M. Srivastava. Recruitment Framework for Participatory Sensing Data Collections. In *International Conference on Pervasive Computing*, 2010.

[3] W. Hsu, D. Dutta, and A. Helmy. CSI: A paradigm for behavior-oriented profile-cast services in mobile networks. *Ad Hoc Networks*, 2011.