Participant Recruitment and Data Collection Framework for Opportunistic Sensing: A Comparative Analysis

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

ABSTRACT

Opportunistic sensing is a novel approach that exploits the sensing capabilities offered by smartphones and users' mobility to sense large scale areas without requiring the deployment of sensors in-situ. In this work, we propose a novel framework for fully distributed, opportunistic sensing which coherently integrates two main components that operate in DTN mode: i. participant recruitment and ii. data collection. We evaluate our approach by considering alternative implementations of the framework, and by measuring their performance via extensive trace-based simulations. Our results show how the performances of the considered protocols vary, depending on the particular scenario, and suggest guidelines for future development of distributed opportunistic sensing systems.

Categories and Subject Descriptors

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

Keywords

Crowdsensing, Opportunistic Sensing, DTN

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 insitu. In opportunistic people-centric 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. For example, an application as the one described in [1], that passively monitors 2G/3G data to assess the quality of mobile Internet from end-terminals, may benefit from the adoption of a distributed data collection model. The users would not need to upload their measurements using the cellular network, therefore saving bandwidth. High participation is crucial for the success of people-centric activities. Because of the low degree of interaction that the user is required to provide, 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. It is therefore critical to devise alternative approaches to opportunistic sensing that are autonomous and fully distributed, and therefore potentially more privacy-preserving. We propose a generic mechanism to perform the *recruitment* and *data collection* activities in a completely distributed fashion. By doing so, we recognize that many different implementations may be applied, and therefore we also recognize the need for a framework to systematically evaluate these implementations. Our major contributions are summarized as follows:

- We introduce a unified distributed recruitment and data collection framework for opportunistic sensing.
- To assess the quality of the compared methods and identify their strengths and shortcomings, a systematic comparative analysis is conducted using a rich set of test suites and scenarios based on real wireless traces.
- The findings indicate guidelines for the use of DTN routing for opportunistic sensing based on sensing mission requirements and the sink model.

2. THE OVERALL FRAMEWORK

Opportunistic sensing activities consist of two phases: *recruitment* and *data collection*. The aim of the *recruitment* phase is to notify an upcoming sensing activity to the nodes that are likely to be able to take part in it. We call this set of nodes *candidate sensing nodes*. Since our focus is opportunistic sensing, where the users are not expected to change their behavior to take part in the sensing activity, some candidate sensing nodes may choose not to participate in the activity. We call the set of nodes that participate in the sensing activity, *sensing nodes*. The aim of the *data collection* phase is to return the sensory data that was collected by the sensing nodes to the organizer of the sensing activity.

MobiCom'13, September 30–October 4, 2013, Miami, Florida, USA
Copyright 2013 ACM 978-1-4503-1999-7/13/09.
http://dx.doi.org/10.1145/2500423.2504573 .

In order to perform distributed recruitment, we propose the use of *recruiter nodes* that advertise the sensing activity, while sensor data will be delivered to a set of *sink nodes*.

2.1 Implementation Details

We used **profile-cast** [3] and **opportunistic geocast** [4] (location-based) for recruitment, and **epidemic routing** [5] (context-oblivious), **prophet** [6] (encounter-based), **spray and wait** [7] (context oblivious) and modified versions of profile-cast and opportunistic geocast for data collection. In this section, we describe the modifications that we applied to these protocols.

2.1.1 Recruitment

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 the original version of profile-cast, where the aim is to reach nodes matching a certain target profile, for recruitment it is necessary to also reach the nodes that match only a *part* of the target profile. It is therefore necessary to change the similarity metric to evaluate the similarity between a subset of the locations and the behavioral profile of the nodes. Upon an encounter with a message holder, the encountered node's profile will be compared to the sets in the power set of the actual target profile and the message will be forwarded if for any of these sets the similarity is above the similarity threshold.

In opportunistic geocast, the expected visiting rate (EVR) is calculated for hexagonal cells that make up the entire routing space for a given geographical region. We modified this metric in such a way that EVR will be calculated for the buildings in the traces, not for the hexagonal cells. However, the two location representations are equivalent, since it is always possible to associate a building with its geographical location. We also devised a geographical dissemination protocol, **geo-dissemination**, which works in a similar way to opportunistic geocast, but nodes may keep a copy of the message and forward it several times.

2.1.2 Data Collection

When the locations visited by the sink are known a-priori, profile-cast may be used for the purpose of data collection; This location information can be used to describe the behavior of the sinks, and therefore included in the recruitment message, and used by the sensing nodes as a target profile for the data messages. Analogously to the recruitment case, data collection aims to reach *any* of the nodes that match the target profile. Because 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 the case.

Opportunistic geocast can also be applied when the locations visited by the sink(s) are known. However, since broadcasting of collected data will result in a high communication overhead, this phase can be omitted and only the ascending gradient phase in the protocol can be applied to reach the sink(s).

When the sink locations are unknown, regular DTN routing protocols such as epidemic routing, prophet or spray and wait may be used without modification, if the sink IDs are known.

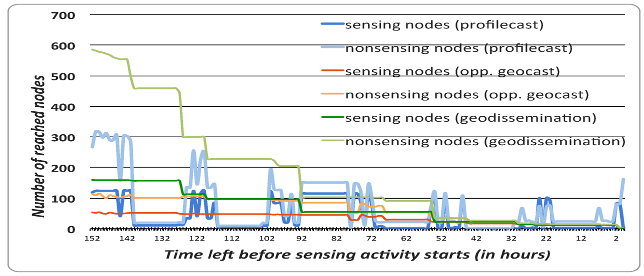

Figure 1: Number of nodes that received the recruitment message (recruited and not recruited) versus time.

3. EVALUATION AND RESULTS

In order to evaluate our approach, we developed a trace-driven simulation in a way analogous to what is described in [3]: we use the extensive WLAN traces that were collected on the USC campus in 2006 [8] and we assume users 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 in the seven most visited locations of the USC campus. A single node, designated both as recruiter and data sink, is deployed one week before the start of the sensing activity, and is either static or continuously moves along a seven waypoint predefined path. The path of the sink is assumed to be included in the recruitment message and this behavior is continuously repeated along the duration of the sensing (from 8am to 8pm).

3.1 Recruitment

The recruiter node is configured to send a recruitment message every hour until the start of the sensing activity. We analyzed the traces and learned that the number of nodes that appear in the sensing area during the activity, and hence should be recruited, is 163.

3.1.1 Mobile sink for recruitment

The recruiter is mobile and travels between the top seven most visited locations. We compare the performances of profile-cast, opportunistic geocast and geo-dissemination for this scenario. While the size of the recruitment message may be negligible, we are still interested in limiting the number of unsolicited messages received by users that will not be able to participate in the sensing activity. Figure 1 shows the number of nodes that received at least one recruitment message for each of these protocols, and the portion of these nodes that will be able to take part in the sensing activity. Geocast recruits the smallest number of nodes (78) with the smallest overhead, and geo-dissemination recruits the highest number of nodes (159) with the highest overhead. Profilecast, on the other hand, achieves a trade-off between these protocols, by recruiting 126 nodes with a moderate overhead.

3.1.2 Static sink for recruitment

The sink is fixed at one the target locations or one of the two non-sensing locations for the duration of the re-

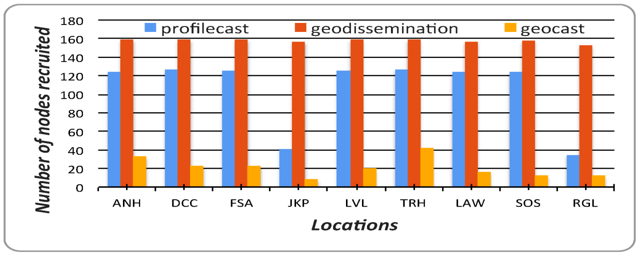

Figure 2: Number of recruited nodes by the protocols for each fixed sink location. All locations except SOS and RGL are target locations from which users are recruited.

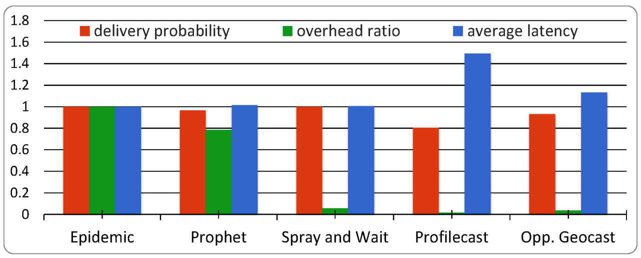

Figure 3: Data collection performance comparison for mobile sink (normalized w.r.t epidemic routing)

cruitment, and stays there until the end of data collection. Figure 2 shows the number of reached nodes by profile-cast, opportunistic geocast and geo-dissemination when the sink is static at a given location.

In case of a static sink, the number of nodes reached by opportunistic geocast decreases significantly. This shows that mobility has a considerable effect on the performance of this protocol. The overhead ratios of the protocols are similar to those in the mobile sink case. Profile-cast offers the best trade-off between the number of reached sensing nodes and reached non-sensing nodes, when the sink is mobile, and it still is a valid option when the sink is static.

3.2 Data Collection

We use the ONE simulator [9] to compare the performances of profile-cast, prophet, epidemic routing, spray and wait and opportunistic geocast DTN routing protocols. We focus on three major performance measures: 1) delivery ratio, 2) average delivery time and 3) overhead ratio.

3.2.1 Mobile sink for data collection

The sink is mobile and travels between the top seven most visited locations. Figure 3 depicts the normalized performance metrics with respect to those of epidemic routing. The highlight of our results is that spray and wait leads to a delivery ratio as high as that of epidemic routing, an overhead ratio that is significantly lower than all protocols except profile-cast and opportunistic geocast, but it beats these two protocols in terms of latency.

3.2.2 Static sink for data collection

The sink is fixed in one of considered locations, and collects data from all the eligible sensing nodes that are present in any of the target locations. Figure 4 shows the normalized average performances (with respect to epidemic routing) of all the nine locations. The performances of all protocols

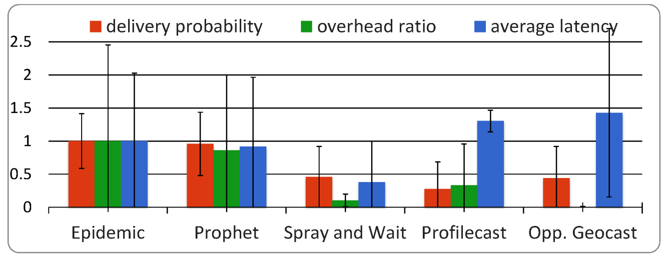

Figure 4: Data collection performance comparison for static sink(normalized w.r.t epidemic routing)

drop significantly for this scenario. The reason for this inefficiency is that the probability of encountering the sink drops drastically, and the messages need to follow a longer route, when it exists. In particular spray and wait's performance drops more significantly than that of other protocols since it blindly replicates messages to a limited number of nodes, without trying to select better forwarders. Figure 4 shows that prophet has the most consistent results with the mobile sink case; it is interesting to observe that although prophet starts collecting encounter information only after the start of the sensing activity, it still achieves a high delivery ratio.

4. CONCLUSION

In this work, we proposed a framework for autonomous and distributed recruitment and data collection in opportunistic sensing. We verified the feasibility of several applicable state-of-the-art DTN routing protocols for both the recruitment and data collection components, and provided a comprehensive analysis of their performances. Our results show that profile-cast is a reliable protocol for participant recruitment due to its comparably lower overhead and higher delivery rate. For the case of data collection, spray and wait outperforms the other protocols for the mobile sink case, although we showed that its performance degrades significantly when the sink only visits a subset of the sensed area, or does not visit the sensed area at all, and therefore there are only few, or no, direct encounters between the sensing nodes and the sink.

References

[1] G. Hourton, G. Del Canto, J. Bustos, and F. Lalanne. Crowd-measuring: Assessing the quality of mobile internet from end-terminals. In *NetGCooP*, 2012.

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

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

[4] Y. Ma and A. Jamalipour. Opportunistic geocast in disruption-tolerant networks. In *IEEE GLOBECOM*, 2011.

[5] A. Vahdat and D. Becker. Epidemic routing for partially-connected ad hoc networks. Technical report, Duke University, 2000.

[6] A. Lindgren, A. Doria, and O. Schelén. Probabilistic routing in intermittently connected networks. *SIGMOBILE Mob. Comput. Commun. Rev.*, 2003.

[7] T. Spyropoulos, K. Psounis, and C. S. Raghavendra. Spray and wait: an efficient routing scheme for intermittently connected mobile networks. In *Proc. WDTN*, 2005.

[8] Mobilib: Community-wide library of mobility and wireless networks measurements. http://nile.cise.ufl.edu/mobilib.

[9] A. Keränen, J. Ott, and T. Kärkkäinen. The ONE Simulator for DTN Protocol Evaluation. In *Proc. SIMUTools*, 2009.