Efficient Data Harvesting in Mobile Sensor Platforms

Uichin Lee[†], Eugenio Magistretti[†] *, Biao Zhou[†], Mario Gerla[†], Paolo Bellavista*, Antonio Corradi*

[†]Department of Computer Science   *Dipartimento di Elettronica, Informatica e Sistemistica
University of California   University of Bologna
Los Angeles, CA 90095   Bologna, Italy 90136
{uclee,zhb,gerla}@cs.ucla.edu, {emagistretti,pbellavista,acorradi}@deis.unibo.it

Abstract

In this paper we study the use of vehicles as sensors in a "vehicular sensor network," a new network paradigm that is critical for gathering valuable information in urban environments. In a vehicular sensor network, each vehicle is responsible for sensing one or more events, routing messages to other vehicles or Infostations and processing sensed data. There are several critical differences between a vehicular sensor network and a "traditional" wireless sensor network, namely: 1) the vehicular network has high computation power; 2) it must provide high storage space; 3) it must have mobile sensor nodes. In addition, due to the sheer amount of data generated, sensed data is carried by each mobile sensor node. Admittedly, retrieving information from mobile sensor nodes is difficult. In this paper, we first identify a set of design choices for building a vehicular sensor network. From this we propose two storage architectures: Content-Addressed Storage (CAS) and Mobility-Assist Storage (MAS). While CAS utilizes Infostations by hashing the key of an event to a specific Infostation, MAS opportunistically disseminates events by "relaying" or sending events only to one's neighbors. CAS is appropriate for time-critical applications and MAS for delay-tolerant applications only if infrastructure is not available. In this paper, we propose data harvesting protocols for CAS and MAS: Infostation-based and mobility-assist data harvesting protocols respectively.

1. Introduction

Wireless sensor networks are likely to be widely deployed in the future because they greatly extend our ability to monitor and control the physical environment from remote locations, e.g. wild habitat monitoring. Recently research communities and automotive companies have been equipping regular vehicles with sensors, thus practically creating a mobile, vehicular-based sensor network (VSN).

In contrast to a traditional static wireless sensor network which consists of a large number of small sensor nodes with low computational, storage and communication capabilities, such limitations no longer apply in a vehicular sensor network. Therefore, we are facing a new paradigm of "mobile" sensor platforms where each vehicle performs event sensing, ad hoc message routing (to other vehicles or to Infostations), processing of sensed data, and filtering. Note that vehicular sensor network shares the same ideas as vehicular ad hoc networks (VANET) [4] in that vehicle-to-vehicle and vehicle-to-infrastructure communications are used, but the ways of collecting, retrieving, and notifying events are generalized as in a traditional sensor network.

A vehicular sensor network can be used in many ways. For example, traffic engineers at the Georgia Institute of Technology have used vehicles to gather traffic information and to better understand how traffic knots. They have pointed out that this will be "one of the most cost-effective ways to reduce congestion, minimize emission, and reduce fuel consumption [3]." Another important application is for the homeland security. To this end, vehicles could be equipped with aural and video devices. Mobile sensor nodes collect a stream of images while cruising down the streets and also detect and notify events to the police. Thus, the police can use such information to investigate terrorist activities.

Unlike "traditional static sensor networks," a mobile sensor network may generate a sheer amount of data depending on types of applications. We assume that a vehicle has enough processing and storage to handle such data. In a traditional static sensor network, sensor data is usually pulled by the data sink using a data-centric protocol such as Directed Diffusion [5]. In a mobile sensor network such as the vehicular network, it is impossible to deliver all the data detected/collected to the sink: first, because just too much is detected by such powerful sensor platforms; second, because the network capacity is too thin because of mobility. We assume that most of the sensed data stays with mobile node. To access data in the mobile storage, we need to de-

sign a protocol that can not only harvest events but also retrieve data from remote storage. However, providing such services is nontrivial in the following reasons. First, because sensor nodes are mobile, we need to keep track of their location to retrieve data from a specific node, thus requiring a location service. Second, for efficient retrieving and searching of large distributed storage, we need an indexing for ease of access.

In this paper we present a set of design choices to solve the aforementioned data harvesting problem in mobile sensor platforms. From this we propose two storage architectures based on the availability of infrastructure and the methods of advertisement: Content-Addressed Storage (CAS) and Mobility-Assist Storage (MAS). CAS utilizes Infostations by hashing the key of an event to a specific Infostation while MAS opportunistically disseminates events by "relaying" or sending events only to one's neighbors. We show that CAS is adequate for time-critical applications, but MAS is used for delay-tolerant applications only if there is no Infostations available. We then propose novel data-harvesting protocols for CAS and MAS. In CAS, Infostations are used to provide all services, such as event retrieval and a location service. But in MAS, a mobile agent harvests opportunistically disseminated data by actively querying its neighbors. [1] In fact, our protocols are based on a cross-layer optimization technique where the application layer of data harvesting is intertwined with routing and location services. The reminder of the paper is organized as follows. In Section 2 we state protocol design spaces of mobile sensor platforms. Section 3 and 4 describe our proposed protocols: Infostation-based and Mobility-assist data harvesting protocols. Section 5 concludes this paper.

2. Protocol Design Issues in Vehicular Sensor Networks

In this section we describe characteristics and assumptions of the vehicular sensor network. We then state the possible design space of vehicular sensor network protocols and propose new storage architectures.

2.1. Characteristics and Assumptions of Vehicular Sensor Networks

We assume that vehicles communicate through a wireless interface, implementing a CSMA/CA MAC layer protocol that provides a RTS/CTS/DATA/ACK handshake sequence for each transmission. To provide a short- to medium-range communication, we also assume that the Dedicated Short Range Communication (DSRC) [9] is supported. This is a new technology operating in the 5.9 GHz

band and achieving a data rate up to 27 Mbps within a range of 1000m. DSRC allows two modes of operations: (1) ad hoc mode characterized by a distributed multi-hop network (vehicle-vehicle) and (2) infrastructure mode characterized by a centralized mobile one-hop network (vehicle-gateway).

Unlike conventional ground sensor networks, vehicular sensor networks are not subject to major limitations such as storage size, computation power, and energy consumption. In a vehicular sensor network, we assume that each vehicle has a relatively fast processor and a large storage device, and has enough energy to operate such devices. In addition, each node is equipped with an onboard sensing device that either generates a small amount of data such as seat occupation or collision detection, or a more complex data such as acoustic sensors and cameras. We assume that vehicles not only have enough storage to handle such streams of data but also enough computational power to process such data to extract features of interest. We assume devices participating in vehicular networks are highly mobile, with a speed up to 30 m/s, but their mobility patterns are predictable due to the constrained movement imposed by the road system and constrained speed imposed by speed limits, traffic conditions and signals. In fact, the mobility of vehicular sensors poses challenges to the communication system. Mobility undermines the reliability of communication and also causes the topology to continuously change.

To summarize, vehicular sensor networks have the following characteristics: (1) no power, computational or storage constraints, (2) high but predictable mobility (e.g., traveling on roads), (3) large-scale deployment (e.g., vehicles in large cities) (4) unreliable communications due to mobility or network partition, and (5) mobile storage that carries vast amounts of generated data.

2.2. Design Space

We first show design space specifically addressing event representation and infrastructure availability. Based on this, we propose two different storage architectures: Content-Addressed Storage (CAS) and Mobility-Assist Storage (MAS).

2.2.1 Event Representation

Since a mobile sensor needs to diffuse events, data representation of sensed events is important. In this paper, for the sake of presentation simplicity, the node can simply diffuse either "data" or "meta-data" of events where "data" is raw information of the event and "meta-data" is short but crucial information of an event. For instance, if a sensor node has taken a picture of a certain location, then it can process the image and extract features of interest. It can further use knowledge bases to automatically generate a meta-data of

---

[1] Here a mobile agent is not a software agent, but a mobile sensor node that is moving and harvesting information from its neighbors.

the image [12]. In fact, in a vehicular sensor network it is impractical to send large images or video clips due to limited bandwidth. Instead of sending full data, a mobile node can send meta-data of events. Thus, a user can query other peers by specifying meta-data of interest. The user then obtains a list of peers where to send data request messages to retrieve data.

2.2.2 Infrastructure

The holistic architecture can be divided depending on the existence of infrastructure. If there are Infostations available, they can be used to provide fixed locations for storing meta-data (harvesting) and a location service (retrieving). Infostations are further classified based on the connectivity between Infostations, but providing connectivity through wired or wireless backbones is expensive and not feasible in many cases, and thus in this paper we focus on Infostations without any connectivity. However, if there is no Infostation available, then we need to rely on epidemic meta-data dissemination techniques to provide such services, i.e. event harvesting and retrieving. In this paper we use a "relaying" technique where data is opportunistically disseminated using node mobility.

2.2.3 Advertisement Architectures

Assuming that we use meta-data, the question is how to advertise meta-data of events. A naïve way is to flood every event, but this is not appropriate for large-scale vehicular sensor networks for the following reasons. First, a flooding based scheme is not scalable. Since every vehicle has some data to report, there will be too many data sources and they will generate too much data. Second, when a network is partitioned, some valuable messages can be lost. In a large scale vehicular sensor network, not only agents but also regular nodes could be located at disconnected locations. Thus, any protocols solely relying on flooding will suffer message loss.

Instead of flooding, we could either publish events to remote locations (remote push) or only to one's neighbors (relaying). In the case of remote push, meta-data of sensed events is stored remotely, e.g., in nodes near a geographic location, by hashing the key of the events. To deal with the high churn rate of mobile nodes, we propose to use Infostations. We call this architecture "Content-Addressed Storage (CAS)." In contrast, in the case of relaying, a mobile node relays meta-data of sensed data to its neighbors, so meta-data is opportunistically disseminated due to mobility. We call this architecture "Mobility-Assist Storage (MAS)." In the following sections, we describe both protocols in detail.
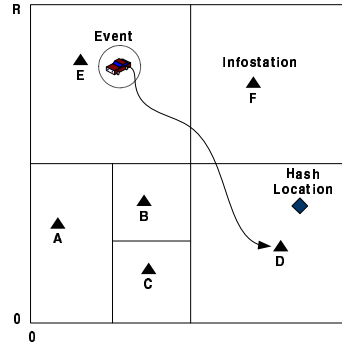

Figure 1. Key Hashing using Infostation Map

3 Infostation-Based Data Harvesting

In this Section we propose a protocol using CAS to minimize event delay latency. To this end, we describe an Infostation deployment strategy which is similar to a common peer-to-peer content distribution architecture [10, 11]. Based on this, we explain how events are efficiently harvested and retrieved.

3.1 Infostation Layout

Infostations can be deployed in a city by considering various facts. One important factor will be traffic conditions; the more vehicles, the more Infostations. When deploying Infostations, we divide 2-D Cartesian space into zones, and for each zone we place one responsible Infostation. Fig. 1 shows an example Infostation layout. For example, Infostation D resides in the area of $\{(R/2, 0), (R, R/2)\}$ and is responsible for storing data which is mapped in that area.

As Infostations are mapped into 2-D Cartesian space, adding or removing an Infostation is not difficult. For example, to add a new Infostation to the zone in which the Infostation A resides, we simply divide the area by half and put the Infostation into the bottom area, $\{(0, 0), (R/4, R/4)\}$. Similarly when we remove an Infostation, the zone will be merged into the larger zone. For example, when we remove Infostation C, then all the data collected in C will be forwarded to B. Note that both mobile nodes and Infostations in the network have the map of Infostations. After any update, the new table must be broadcast. Since Infostations are static, such updates rarely occur, and thus the costs of such updates is very limited.

3.2 Event Harvesting

To store events we simply assume that meta-data contain a key field. Then, we determine a geographical location, by hashing the key through a couple of hash functions that output x and y coordinates. By looking up the Infostation
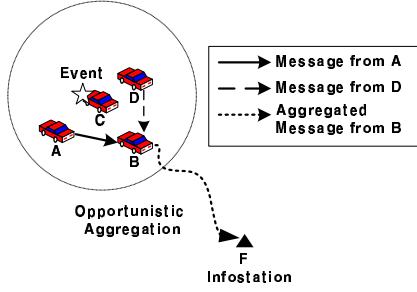
Figure 2. Opportunistic Data Aggregation

Figure 3. Event Notification

map, a sensor node can find the Infostation responsible for that location. From this, it can send data to the destination Infostation. For example, in Fig. 1, a newly detected event is mapped into the zone which is under control of Infostation D. It forwards the message to D using a typical geo-routing scheme [6]. In the case of event notification, we need to use a pub-sub system in Infostations. Those who are interested in some events must register with Infostations to receive the events.

When detecting an event, sensor nodes near the source of the event tend to detect the same event at the same time. For example, in Fig. 2 nodes A and B detect the same abnormal feature of node C and then want to publish the data. In this case, the feature is mapped into Infostation F. Multiple copies of the event will be forwarded to F using a geo-routing scheme. Thus, when the messages are forwarded to F, they will primarily travel the same path. This allows us to opportunistically aggregate multiple copies of the same event, thus minimizing unnecessary data transfer to the target Infostation.

As meta-data is stored in Infostations, Infostations become distributed indexes. For example, the agent sends a query and then gets a list of peers that actually have the data of interest. Then, it can send data request queries to those peers. Another approach is that an Infostation processes a query and then sends the result back to the user. This incurs computational and communication overhead, but if the data accessed shows temporal correlation, i.e., is repeatedly requested over a period of time, then this actually reduces overall communication costs. Note that when a network is partitioned and intermediate nodes are not able to forward messages to Infostations, we use "relaying" techniques which we will elaborate in the next section.

3.3 Data Access

In order to support the access to the actual data, our cross-layer solution includes a routing protocol. Generally, in vehicular sensor networks packets are routed according to geographic forwarding schemes. However, a prerequisite of geographic routing is a location service which allows a
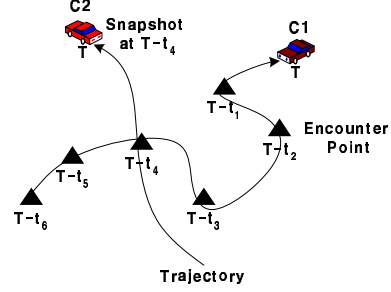
source node to obtain the location of a destination before data traffic follows. In contrast to this approach, we exploit an hybrid approach that intertwines a location service and routing using node encounter history [1, 2]. The intuition behind such schemes is that node-encounter history or last-encounter information generally provides a rough yet useful estimate of the current network topology. It is interesting to note that we can publish last-encounter information in the same way that we do regular features. For example, in Fig. 2 node C's last-encounter information can be published by its nearby nodes A, B, and D. This information will be delivered to Infostation F, and in the same way we can opportunistically aggregate the information. Since last-encounter information is based on time, we set the maximum threshold that we can aggregate between two arrivals. Let t be the aggregation window. Any encounter information that arrives within period t will be simply dropped and will not be forwarded. An important fact is that when the data packet is sent, it is originally set as a stale location due to the delay introduced by the update aggregation. But when the packet travels near the area, as in the dotted circle in Fig. 2, nearby nodes have fresh last-encounter information of the destination and thus, by refining estimates, the packet can be delivered to the destination.

4 Mobility-Assist Data Harvesting

In [7, 13] relaying is used to deliver a message to the destination in case of temporal disconnection. In this paper we further extend this idea and use it for both data *harvesting* and remote *data access*.

A mobile sensor node meets other sensor nodes when they are within its communication range. At that moment the sensor node advertises the data collected over a fixed period of time to its neighbors. Each neighbor node will receive a snapshot of collected data by the sender sensor node. In this way, the data is opportunistically diffused into the network. Finally, a mobile agent harvests the collected data by proactively querying its neighbors. Note that a mobile agent also can detect an event in the same way. For

example, in Fig. 3 there are two sensor nodes, C1 and C2. While sensor node C1 travels, it gathers information. A black triangle sign denotes that C1 has encountered other sensor nodes. For ease of explanation, we assume that there is only a single encounter, but in reality any nodes within C1's communication range are considered encounters. In this example, assuming that the current time is T, at time $T - t_4$, C2 encounters C1. At that moment, C2 can receive all the meta-data collected over a fixed period of time by C1. Then the agent can later query C2 and get snapshots from C1 and C2 at time $T - t_4$.

To access data, as in "Infostation-Based Data Harvesting," we use last-encounter-based routing protocols [1, 2]. To this end, last encounter information is piggy-backed when mobile nodes are advertising meta-data of events. An agent queries its neighbors to retrieve meta-data collected which also contains information of data source. Sensor nodes are continuously receiving advertisements, and update last encounter information of encountered nodes. If the agent finds meta-data of interest later on, it can send a data request packet to the target using last-encounter-based routing protocols.

Note that this idea is related to data MULEs [8] in static sensor networks which pick up data from sensors when in close range, buffer it, and drop off the data to wired access points. However, our scheme uses mobility-based opportunistic dissemination to further expedite data harvesting.

5 Conclusion

In this paper we investigated the space of protocol design alternatives for vehicular sensor networks. Based on the availability of infrastructure and the methods of advertisement, we proposed two different storage architectures: Content-Addressed Storage (CAS) and Mobility-Assist Storage (MAS). In CAS, meta-data of sensed events are stored remotely, i.e., in a geographic location by hashing the key of the events. But in MAS, a mobile node relays meta-data of sensed data to its neighbors, so meta-data is opportunistically disseminated due to mobility. For CAS, we proposed an Infostation-based data harvesting protocol where the network is divided into zones to deploy Infostations, and geographic hashing is used to publish and retrieve information. For MAS, we proposed a mobility-assist data harvesting protocol where relaying is used to disseminate sensed data, and routing and a location service are also provided to access events of interest. The reader will observe that there are applications which are delay tolerant, but with finite delay constraints. In this case, if both search modes are available, it will be of interest to investigate the tradeoffs of using one or the other scheme. The choice will depend on various parameters, including mobility patterns, data traffic load, etc.

References

[1] H. Dubois-Ferriere, M. Grossglauser, and M. Vetterli. Age matters: efficient route discovery in mobile ad hoc networks using encounter ages. In *ACM MOBIHOC'03*, pages 257–266, 2003.

[2] M. Grossglauser and M. Vetterli. Locating Nodes with EASE: Mobility Diffusion of Last Encounters in Ad Hoc Networks. In *IEEE INFOCOM'03*, March 2003.

[3] E. Guizzo. Network of Traffic Spies Built Into Cars in Atlanta. *IEEE Spectrum*, April 2004.

[4] H. Hartenstein, H. B. Fubler, and A. Festag. FleetNet - Vehicular Ad Hoc Network for Inter-Vehicle Communications. In *ITS'03*, November 2003.

[5] C. Intanagonwiwat, R. Govindan, and D. Estrin. Directed diffusion: a scalable and robust communication paradigm for sensor networks. In *ACM MOBICOM'00*, pages 56–67, 2000.

[6] B. Karp and H. T. Kung. GPSR: greedy perimeter stateless routing for wireless networks. In *ACM MOBICOM'00*, pages 243–254, 2000.

[7] D. Nain, N. Petigara, and H. Balakrishnan. Integrated Routing and Storage for Messaging Applications in Mobile Ad Hoc Networks. In *WiOpt'03*, Sophia-Antipolis, France, March 2003.

[8] S. J. Rahul C. Shah, Sumit Roy and W. Brunette. Data MULEs: Modeling a three-tier architecture for sparse sensor networks. *Elsevier Ad Hoc Networks Journal*, 2003.

[9] S. J. Rahul C. Shah, Sumit Roy and W. Brunette. Standard Specification for Telecommunications and Information Exchange Between Roadside and Vehicle Systems - 5 GHz Band Dedicated Short Range Communications (DSRC) Medium Access Control (MAC) and Physical Layer (PHY) Specifications. *ASTM*, September 2003.

[10] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker. A Scalable Content-Addressable Network. In *ACM SIGCOMM'01*, 2001.

[11] S. Ratnasamy, B. Karp, L. Yin, F. Yu, D. Estrin, R. Govindan, and S. Shenker. GHT: A Geographic Hash Table for Data-Centric Storage. In *WSNA'02*, Atlanta, Georgia, USA, September 2002.

[12] R. Sarvas, E. Herrarte, A. Wilhelm, and M. Davis. Metadata Creation System for Mobile Images. In *ACM MOBISYS'04*, Boston, MA, USA, 2004.

[13] W. Zhao, M. Ammar, and E. Zegura. A Message Ferrying Approach for Data Delivery in Sparse Mobile Ad Hoc Networks. In *ACM MOBIHOC'04*, Tokyo, Japan, May 2004.