Data Muling with Mobile Phones for Sensornets*

Unkyu Park
Information Sciences Institute
University of Southern California
Marina del Rey, California, USA
unkyu@isi.edu

John Heidemann
Information Sciences Institute
University of Southern California
Marina del Rey, California, USA
johnh@isi.edu

Abstract

Sensors are all around us, in buildings, vehicles and public places, from commodity thermostats to custom sensornets. Yet today these sensors are often disconnected from the world, either because they are distant from infrastructure, and wide-area networking (by 3G cellular, satellite, or other approaches) is too expensive to justify. Data muling makes communication cost-effective by leveraging short-range wireless and mobility, perhaps by zebras, buses or farmworkers. In this paper we propose that *human-carried mobile phones* can serve as data mules for sensornet deployments, exploiting ubiquity of mobile phones and human mobility to bring low-cost communication to sensors. We use two mobile phone datasets to show that Bluetooth can serve as a viable muling network, and humans already see many potential sensors regularly. We have implemented a mobile-phone-based data muling system, and used it in four sensornet deployments totaling ten months operation. We find that muling can be the only cost-effective option for rural deployments, where it is critical to monitoring remote sensor networks. We also show opportunistic mobility can collect data without any extra effort in residential and office environments. Finally, we systematically evaluate our deployments to understand how contact duration and data size interact, and to evaluate the effect of muling on phone batteries.

Categories and Subject Descriptors: C.2.2 [Computer-Communication Networks]: Network Protocols

General Terms: Design, Algorithms, Performance, Experimentation, Measurement

Keywords: data muling, wireless sensor networks, mobile telephones, human mobility

* This research is partially supported by by the National Science Foundation (NSF) under grant number CNS-0626702, "Sensor-Internet Sharing and Search" and by CiSoft (Center for Interactive Smart Oilfield Technologies), a Center of Research Excellence and Academic Training and a joint venture between the University of Southern California and Chevron Corporation.

SenSys'11, November 1–4, 2011, Seattle, WA, USA.

1 Introduction

Our world is increasingly instrumented, with useful electronic information everywhere: office buildings have a thermostat in many rooms and motion detectors at doors; automobiles have dozens of sensors monitoring the engine, passengers, and outside environment. Today these devices serve their original, dedicated purposes—yet what if their data were made available through low-cost communication?

Sensor networks begin with the premise that sensors can communicate. SCADA systems have been used in industry for decades, and industrial SCADA systems and sensornets are increasingly sophisticated, but often price limits their use to high value applications. Similarly, scientific uses of sensornets are growing, but connectivity to remote locations is expensive, limiting less fortunate scientists to datalogging without experiment supervision. How can we make wide-area communication more accessible to sensornets?

Wi-Fi, cellular telephone data networks, and satellite communication make wide-area communication commonplace. But these approaches do not solve these problems because they are often too expensive or unavailable where needed. Satellite data is often prohibitively expensive. Cellular seems more affordable, but monthly fees for data service are often too high for many applications. (Even services such as the Amazon Kindle that are free to the user embed data fees in the cost of content, discouraging high-bandwidth content [13].) And if cost is not considered, *coverage* remains a problem. Wi-Fi is free, but coverage can be spotty even in urban areas. Satellite coverage requires line-of-sight to the sky. Cellular coverage is generally good, but all providers have dead spots. Humans work around coverage problems by moving, but that option is not available to fixed sensors. Some sensornets use a local mesh to get to wide-area connectivity, but that approach greatly adds to complexity and failure cases for simple, small deployments.

In this paper we describe our system where *human-carried mobile phones* serve as data mules for sensornet deployments, exploiting ubiquity of mobile phones and human mobility to bring low-cost communication to sensors. Many other groups have explored the idea of data muling for sensor networks [4, 9, 20, 21, 25, 28, 29, 35], and some have proposed human mobility for communication in remote areas [18], in disasters [12] or from cars [19].

Our work makes three contributions beyond the prior work. First, we bring together sensing and the mobile phone as a data mule in our implementation of muling in an off-the-shelf mobile phone (§3). We also describe several applications where muling's cost reduction makes new applications viable or current applications easier to justify (§2.2).

Second and more importantly, we have used our phone-based mule in four sensornet deployments: a field deployment for subsidence detection in an oilfield, two different testbeds emulating that application but in an urban area, and an office-based person-detector (§4). To our knowledge, we are the first to use human-carried mobile phones to collect data from real sensor deployments in remote areas. Together these ten months of deployment experience help us understand what makes data muling work in practice. We show that muling provides essential feedback for experimental deployments in remote areas (§6.1.2), halving the time outages of experimental hardware were unknown, from 60 sensor-days to 27. These deployments drive evaluation of design trade-offs such as use of Bluetooth or Wi-Fi for sensor-to-mule communication (§6.3). We also evaluate when energy consumption is a limiting factor (§6.4).

Our last contribution is to explore how human mobility patterns affect the potential of data muling. We examine two datasets of mobile phone contacts to show that humans see many potential sensors (§5.1), and some of these regularly (§5.2). Yet our deployments show that *intentional* mobility is often required when coverage of specific sensors is required, at least with ranges typical for Bluetooth (§6.1). We show the importance of the human's loiter time to effective muling (§6.2), and that that opportunistic mobility works best in our office deployment where sensors are dense and loiter times are long (§6.1.4). We also consider an alternative communication choice with long range radio and fast data rate, showing that it benefits the data muling and make opportunistic muling even more practical (§6.3).

2 Motivation, Applications, and Challenges

We next describe why mobile phones make good mules, and show applications that motivate phone-based muling.

2.1 Why Data Muling using Mobile Phones?

Both sensors and people with mobile phones are all around us, but the cost of wide-area networking for cheap sensors is often prohibitive. We suggest that mobile phones can bridge this connectivity gap through data muling.

Mobile phones are attractive as data mules for three reasons. First, mobile phones are truly ubiquitous. In fact there are approximately 4.6 billion mobile phone users worldwide estimated by the International Telecommunication Union. That means 68% of the worlds population already carry mobile phones all the time. And mobile phones are widely used in the developing world where the need for data muling is greatest since other forms of wide-area communication are often limited. Although currently most phones in the developing world are feature phones with limited extensibility, in principle even these telephones could support muling, and we expect phone capabilities to grow.

Second, smart-phones today are powerful, general purpose computing platforms. They already include energy-conserving, short-range radio networks like Bluetooth, and with wide-area Internet connectivity through 3G and now 4G telephony. We show later (§6.4) that we can use these networks for muling with minimal additional energy cost.

Third, mobile phones are already carried by humans every day, so muling can piggy-back on this mobility for free.

Finally, the large display and physical or virtual keyboard of modern smartphones provide a friendly interface to sensors. Many embedded sensors lack a sophisticated interface or on-site control, and as sensors become smaller, lower-power, and cheaper, user interfaces become impossible to provide. We anticipate the mobile phone can be used a unified interface to various sensors.

2.2 Motivating Data Mule Applications

We next describe five applications that are good matches for data muling—each needs sensing, can tolerate variable and sometimes large delays in data retrieval, and are in areas with human mobility and without reliable or cheap wide-area networking. With near-ubiquity of mobile phone coverage, it may seem that wireless coverage should be always available. However, we show that in our Subsidence/Oilfield application, cellular data coverage was so poor as to be unusable (§4.2), and even when cellular coverage is good, its price or energy draw is often too high to justify its use.

Assisted-reporting Garbage Bins: Today garbage bins in in national parks and urban public spaces are often serviced with a fixed, periodic schedule. A fixed schedule works poorly when bursts of use fill bins unexpectedly, or underuse results in needless trips for servicing. We expect that need-based servicing can reduce maintenance costs and improve citizen satisfaction.

Garbage bin monitoring is ideal for data muling because they are often sparsely deployed in remote areas, yet they serve humans carrying mobile phones [32].

Habitat Monitoring: Habitat monitoring has been studied by many sensornet researchers [7, 34, 30]. Several deployments to-date have used long-range wireless or satellite connections to relay observations to researchers' institutions, but expensive and custom networking may be challenging to justify for smaller habitat monitoring projects. We suggest that data muling can lower the cost and technical requirements for habitat monitoring by exploiting the mobility of humans as they travel to the target habitat then back to urban areas with inexpensive networking. Even if some habitats lack regular hikers, opportunities for data muling may be sufficient with park rangers or scientists.

Car Blackboxes: Prior work has brought sensing to vehicles [5, 23, 19]. Vehicles have much information to provide, from gas mileage to details about engine performance and safety. Applications may mine these data archives to suggest needed maintenance or give feedback to car manufacturers. While projects such as CarTel have shown one can exploit opportunistic Wi-Fi connections [19], increased security concerns mean that availability of open Wi-Fi networks can be inconsistent, and owners may not be motivated to extend their home Wi-Fi network to parking areas. Our data muling system can replace Wi-Fi connections with Bluetooth connections to the driver's mobile phone (CarTel suggested, but did not explore, this possibility [19]).

Personal Energy Monitoring: Personal energy conservation is of growing interest, partly because simple knowledge of energy consumption allows individuals to reduce

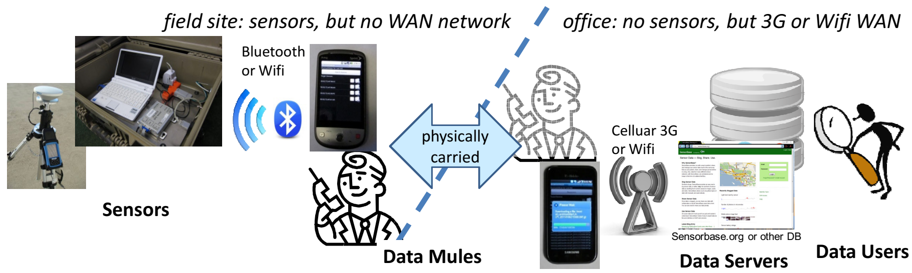


Figure 1. An overview of our data muling system.

consumption by 5–10% [17]. Several projects have begun instrumenting home power usage, including Google PowerMeter [16] and MS Hohm [24]; others such as LoCal [22] are exploring smart grids that negotiate electricity usage between suppliers and consumers.

While these approaches provide a region- or home-centric view of energy use, we suggest that data muling can provide a *personal* model of energy use—capturing use at work and in public spaces as well as at home. A personal view of energy consumption requires harvesting data from sensors in our environment. We suggest that data muling can provide this data: devices that consume energy can have Bluetooth-based energy sensors, and people's mobile phones monitor devices around each individual. Data muling in this case is less about connectivity than about discovery and recording the correct information. While energy data may be too sensitive to post publicly on the Internet, the physical proximity provided by muling may reduce those concerns, since proximity implies some relationship with the provider of the data being collected.

Hiking Water Quality: Hikers often face questions about water quality at remote fountains or natural springs or streams. Governments or informed hikers may inspect water quality, but today there is little way to retrieve water data from remote locations, or to share that data with other hikers. Remote locations have little data infrastructure, and although important to health, the economic value of monitoring hiking water cannot support satellite data. As shown by CenWits [18], data muling can exploit hikers' mobility to deliver the information of water quality cost-effectively to those who care about it. We suggest that the mobile phones likely already carried by the hikers are an ideal data mule today.

2.3 Data Muling Challenges

Prior work has shown the principles of data muling: relaying data between nodes upon rendezvous, exploiting random mobility [21] or expected mobility patterns [9, 25]. We build on this prior work and answer several new challenges:

How effective are current mobile phones as data mules? We examine mobile phone hardware, evaluating Bluetooth, Wi-Fi, and 3G cellular technologies for networking. We consider evaluate energy consumption and the unique constraints of mobile telephones, with typical daily charging and an important requirement to never run out of power.

Is data muling feasible for traditional sensornet applications? We explore four sensornet deployments retrieving different kinds of data to understand how well data muling works in practice. Important in our study is understanding how mobile phones work in these environments, how real-world data sizes and latency requirements affect muling, and if muling is effective for these applications. We also compare loiter time, from human mobility, and muling time, driven by application requirements.

How well does human mobility support data muling? The success of muling depends on the mobility pattern. Ideally we would like muling to be "free", leveraging human mobility. We study new and existing datasets of human mobility, and then evaluate intentional and opportunistic mobility in our deployments.

3 Design of Our Data Muling System

The three components of our data muling system are shown in Figure 1: sensors, mules, and gateways. *Sensors* generate data, store it locally, and periodically listen for a mule to come by. Humans carry *data mules* around sensors; mules periodically scan for sensors and automatically fetch any new data they see. Humans may carry the data mules with *intentional*, planned meetings with sensors, or they may simply go about their daily business and rendezvous happens *opportunistically*. Mules and sensors communicate over a short-range, low-power, wireless network; we use Bluetooth and 802.11, while 802.15.4 is third option.

Data mules only store data temporarily. They relay their data to Internet-based *data servers* for user analysis. Our implementation uses either 3G cellular networks or 802.11 for mule-to-server communication, based on the preference of the mule carrier and convenience of 802.11. Muling allows two important advantages relative to traditional sensor networks: first, the sensornet need not have a dedicated Internet gateway with corresponding need for extra power and cost. Bluetooth is inexpensive enough we use it on all sensors, and the mule amortizes the cost of WAN communication over all sensors. Second, when sensors originate data in an area with poor WAN coverage, a human can take a mule from this data-rich/network-poor area to another area with good or free WAN coverage. Both of these advantages motivated our adoption of data muling in our four deployments.

Next we describe each component of our data muling system in more detail.

3.1 Sensors

Details of sensor hardware depends on the specific application. In addition, muling requires some low-power, short-range wireless networking protocol.

In our deployments our sensors use embedded PCs with Bluetooth as the short-range wireless protocol. We use Bluetooth primarily because it is available on the mule, but as a commodity protocol it is widely available at very low cost. Many sensors today include Bluetooth, including weather stations and automobile accessories. Our custom-built sensors did not include built-in Bluetooth, so we added it with an inexpensive (around US$10) USB adapter. In principle, other short-range, low-power wireless networking protocols could replace Bluetooth. One promising protocol is 802.15.4 because of its low power.

We also use 802.11 as the sensor-to-mule network in §6.3 to evaluate how much larger bitrates change the experience. While not low power, its much higher speed reduces loiter time (§6.2). Ultimately we look for Bluetooth 3.0 to combine low-power rendezvous and 802.11-speed communication.

As wide-area network hardware costs fall, Wi-Fi and cellular model hardware may be cheap enough that it is in sensors "for free". Muling still has a place because sensing often forces sensor placement into specific locations with poor *local* coverage, in spite of generally good WAN networks. (We encountered this specific problem in our Subsidence/Oilfield application as discussed in §4.2.) Low-power, short-range networks can also reduce energy consumption compared to WAN communications.

Our muling scheme is agnostic to the type of sensor. We use two sensors in our deployments. Two deployments use GPS sensors that capture files that are 1–2 MB in size (after compression) every two hours, generating 12–24 MB/day. The third deployment tracks people via their carried Bluetooth devices; this sensor generates very small datasets, typically less than 800 kB per day.

3.2 Data Mules

Our data mules are mobile phones. We use mobile phones to exploit these intelligent devices carried by almost everyone today. We next describe our hardware and software choices for our mules.

We currently use four different Android-based smartphones as our mules: the HTC Hero, HTC Touch, Samsung Galaxy S, and HTC EVO. These platforms were chosen because of their availability and suitability for an individuals personal mobile phone. These platforms also drive our choice of Bluetooth as the sensor-to-mule wireless protocol, so that no hardware modifications to the mule is required. We also considered but have not yet implemented a mule based on an embedded PC with 802.15.4 support so we could mule from motes and other embedded sensors. We currently require support for Android 2.0 for Bluetooth scanning supported only since that release. In principle our muling software should port to other smartphones such as the Apple iPhone or phones based on Windows or SymbianOS.

On the mobile phone we run our custom muling service. The muling application runs as a background service, periodically scanning for neighboring Bluetooth devices to determine if they are sensors. We use a default scan interval of 2 minutes, although configured from 1 to 10 minutes to tradeoff detection speed against energy consumption (we look at energy consumption in §6.4).

The mule has a list of known sensors, and when one is within range it connects to the sensor and retrieve and fresh data. In addition to gathering data from the sensor, the mule pushes a delivery report of what data has been delivered to the Internet, allowing the sensor to garbage collect data. Mules can handle many sensors, limited only by battery and storage. Modern phones have storage for thousands of large data items, and we disable muling when the battery is low, so our system can adapt to even large numbers of sensors.

We do not coordinate multiple mules; each operates separately and the first to encounter a sensor retrieves all pending information. We currently assume mules are trustworthy relays. Sensors could encrypt data and send copies through multiple mules in less trusting environments, or disable muling if there are far more mules than needed.

Although it is not essential for successful muling, our application includes a user interface that reports what sensors are in range and what data has been collected on the mule. The interface can be used for other purposes well, if a sensor needs servicing (perhaps battery replacement or sensor cleaning), the mule could request assistance from the carrier.

3.3 Data Servers on the Internet

We expect that all data is ultimate hosted on servers on the Internet (as is the case in nearly all operational sensornets). Our current implementation uses two different storage servers. We use a webserver with an off-the-shelf uploading extension to support our two data-intensive deployments. Subsidence/Oilfield and Subsidence/Urban(§4) each generates 84–168 MB per week. For our third application (People/ISI), we store data in Sensorbase.org [10], a sensor data sharing platform built on Apache and MySql. Sensorbase also includes support for managing and sharing sensor data, and allowing users to query and interact with stored data.

3.4 Design Alternatives

We next briefly discuss alternatives that we considered, and our reasons for not deploying them.

Mote-based Mules: We considered both motes and mobile phones as mules. We use mobile-phones as mules because they are already carried and have excellent form-factor and battery life due to commoditization.

We prototyped a data muling system using Mica2 motes to understand the potential 802.15.4-like mote-to-mule communication. Mica-class-devices are attractive sensors because of their proven success at long-term, energy conscious operation, combined with their easy customizability.

Mule-side support for 802.15.4 is a barrier, however, since no phones support 802.15.4. We considered having users carry a mote as the mule, but use of a second device just for muling is a significant burden. While we believe many users would run muling if it had no impact on their use of existing devices, we expect that few would add a new device to their daily lives for this purpose. As a secondary concern, standard motes have less than 1 MB flash storage, too little for our Subsidence applications. While future phones support new radios such as Bluetooth 3.0, Near Field Communication, or perhaps software-defined radios, causal muling requires sensors to conform to consumer standards.

Multi-hop communication between sensors: Our current system assumes all sensors can directly communicate with the mule. With many multi-hop communication (mesh) protocols for sensornets, we could easily employ multi-hop communication between sensors.

A mesh network between sensors is of interest only when sensors are clustered, physically close to each other. In that case, the sensors could preemptively push data to a designated collection point, or the appearance of a mule could prompt the sensors to gather their data on-the-fly. Either way, mesh communication increases opportunities to mule data. We believe the greatest advantage of multi-hop networking among sensors is that it can extend the energy-efficient network to locations that are difficult for the mule to reach. A second advantage may be coordinated sleeping among sensors to conserve energy (as explored for other purposes [36, 11, 27]). However, the cost of a sensor mesh is much greater complexity, to insure a connected mesh and to manage resource usage at a designated collection point. We did not employ sensor-to-sensor communication because our deployments have one or two clusters of sensors, so multi-hop was not necessary at each site, where all sensors could be reached with one visit, nor possible between sites.

Multi-hop communication between mules: We also considered and rejected multi-hop communication between mules, as in prior sensornet muling deployments [21, 18] or between mobile phones during disasters [12]. Given the wide-spread availability of cellular data connectivity our goal is to mule data out of dead spots with poor network coverage but sensing interest, and to amortize the cost of the data connection among many sensors. Mule-to-mule communication does not help either of those goals, so we do not consider mule-to-mule connectivity.

4 Case Studies: Human Mobility and Sensornet Deployments

We employ six datasets in this paper to understand data muling potential and practice. As shown in Table 1, the first two, Mobility/MIT and Mobility/ISI, are passive observations taken from mobile phones, while the others are four different sensornet deployments we undertook to study data muling. These studies show a wide range of muling scenarios, with both intentional and opportunistic mobility patterns; different numbers of mules and sensors; weekly, daily, or more frequent mule visits; and small and large data sizes (bytes to megabytes) .

4.1 Observations of Human Mobility

We began our work with a public dataset about mobile phone mobility, then conducted additional experiments to improve precision.

The *Mobility/MIT* trace is from the Reality Mining project, where they collected mobile phone activities from 100 mobile phones for 9 months [14]. (Their full dataset runs 18 months, but we use the 9 months starting in Sept. 2004 that they identify as their active data collection period.) Their public dataset includes rich information including calls, location, and Bluetooth contacts. We study their Bluetooth contact information to estimate regularity in human mobility to show the potential for data muling in §5.

The Mobility/MIT dataset scans for neighbors at 5-minute intervals. To observe brief connectivity, we carried out a smaller Bluetooth survey with mobile phones with the *Mobility/ISI* dataset. We scan every two minutes, a period chosen to balance battery life and frequency of detection. Each scan takes about 15 seconds. We use this additional dataset to update the prior dataset and to better understand opportunities for muling in §6.

4.2 Sensornet Deployments

We have employed our data muling system with four sensornet deployments.

The *Subsidence/Oilfield* deployment involves two pairs of GPS units observing subsidence in a production oilfield. The project carried out multiple deployments over several years and adopted data muling for the most recent 4-month deployment from Oct. 2010 to Jan. 2011 out of necessity. The experimental hardware required close monitoring to insure correct operation. Unfortunately, the industrial field wireless network was not ubiquitous, nor were we allowed to access. Early deployments used 2G and 3G cellular modems for data, but we were unable to get consistent cellular coverage for more than a few days. We then tried manual data retrieval by swapping flash cards, but month-long intervals and difficultly swapping cards (travel to site, open locked box, halt machine, etc.) made "sneakernet" untenable. We therefore deployed to data muling.

Unlike the mobility datasets, muling for Subsidence/Oilfield is *intentional:* with only one mule and a large oilfield, field personnel would explicitly drive to each approximately weekly. Although in principle one could have swapped memory cards, data muling greatly simplified data retrieval, since it requires only wireless connection, personnel need only drive nearby, park, and push a button on the smartphone.

The *Subsidence/Urban* dataset uses the same equipment as Subsidence/Oilfield. However in this case, subsidence is part of a controlled experiment and the site is at a residence in an urban area. As a residence, we were able to frequently mule (except for travel and operator error). Because the sensors are out of Bluetooth range of the residence, most muling was again intentional, however in §6.1.3 we show that in many cases, normal movement was near enough to the sensors provide opportunistic muling as well.

Finally, the *People/ISI* dataset is designed to provide pure opportunistic muling in an office environment, and also emphasizes small data sizes, with each report tens of bytes rather than megabytes. We deployed sensors in four locations: two offices and a break room at ISI, and the home of a researcher. Each sensor tracked nearby people, as determined by scanning for Bluetooth contacts; each location had a number of visitors. Muling happened only opportunistically, as the operator carried his mobile phone as part of daily use. In §6.1 we use this dataset to evaluate the effectiveness of casual muling in daily life.

5 Evaluating the Potential for Data Muling

We next consider opportunities for data muling. We first look at how pervasive short-range wireless sensors could be, and how regularly humans visit them. For both of these

Dataset	Goal	Description	Mules	Mobility	Data Size	Sensors	Start	Duration
Mobility/MIT [14]	observation	Bluetooth scanning log (5-minute interval)	100	opportunistic	—	815*	Sept. 2004	9 months
Mobility/ISI	observation	Bluetooth scanning log (2-minute interval)	3	opportunistic	—	226*	May. 2010	12
Subsidence/Oilfield	deployment	Oilfield subsidence monitoring	1	intentional	~2 MB	4	Oct. 2010	3
Subsidence/Urban-BT	deployment	Subsidence monitoring in urban area (Bluetooth)	1	mostly intentional	~2 MB	2	Nov. 2010	4
Subsidence/Urban-Wi-Fi	deployment	Subsidence monitoring in urban area (802.11)	1	opportunistic	~2 MB	2	Jun. 2011	1
People/ISI	deployment	Person monitoring in office area	1	opportunistic	~10 kB	4	Feb. 2011	2

Table 1. Datasets considered in this paper: observations and sensornet deployments. (* indicates sensor stand-ins)

questions we consider our two observational datasets (Mobility/MIT and Mobility/ISI) and use Bluetooth devices as a stand-in for sensors, as described below. Then in the next section we revisit these questions with our system in practice in four deployments.

5.1 How Many Potential Sensors Around Us?

Data muling presumes that short-range wireless sensors are pervasive and available for muling. Today wireless sensors clearly are *not* everywhere, although *each of the pieces exists*: sensors, wireless communication, and sensor networks. Sensors are deployed and operating in almost everywhere we go: thermostats, motion detector at door, smoke detectors, power meters and water meters, several sensors in each mobile phone, hundreds of sensors in automobiles, and cameras and pressure sensors on streets. Yet today these sensors often stand-alone or are used only in specific application “stovepipe”. Wireless communication is everywhere as well, with Wi-Fi, Bluetooth, and 3G and now 4G mobile phone data. And there have been a number of long-term sensor network deployments as well. Yet we conjecture that there is a bootstrapping problem: there is no muling today because there are few public, wireless sensors, and yet there are few such sensors because there is no muling.

To break this deadlock we first wish to characterize how wireless sensors *might* operate. Here we consider the effectiveness of short-range wireless communication to answer how many sensors might we see; in the next section we evaluate how regularly humans would see those sensors. In both cases, we use *stationary Bluetooth devices* as a *stand-in* for wireless sensors. We target only *stationary* devices because we expect most environmental sensors to be stationary. We consider Bluetooth because it is cheap enough (current Bluetooth chipsets add only pennies to the cost of a device), and low-enough power that it is a plausible technology for a public wireless sensor.

We recognize that our sensor stand-ins are an imperfect prediction of where actual networked sensors may be placed in the future. We use them because they allow the study of real data about human mobility and Bluetooth propagation, although against stand-in, approximated sensors. We therefore augment this “what-if” analysis with four real sensornet deployments in §6.1, showing that Bluetooth connectivity can work, but Wi-Fi’s greater range is helpful.

The Mobility/MIT dataset dataset has 25,687 unique Bluetooth devices observed by 100 mobile phones over 9 months. In Mobility/ISI dataset, we observed 12,954 unique devices from a single phone. We attribute the count in Mobility/ISI to its collection 7 years after Mobility/MIT and increasing use of Bluetooth.

	Mobility/MIT	Mobility/ISI
devices encountered	25,687	12,954
mobile devices	23,814	12,699
stationary devices	815	226
unknown	1029	29
start	Sept. 2004	May. 2010
duration	9 months	12

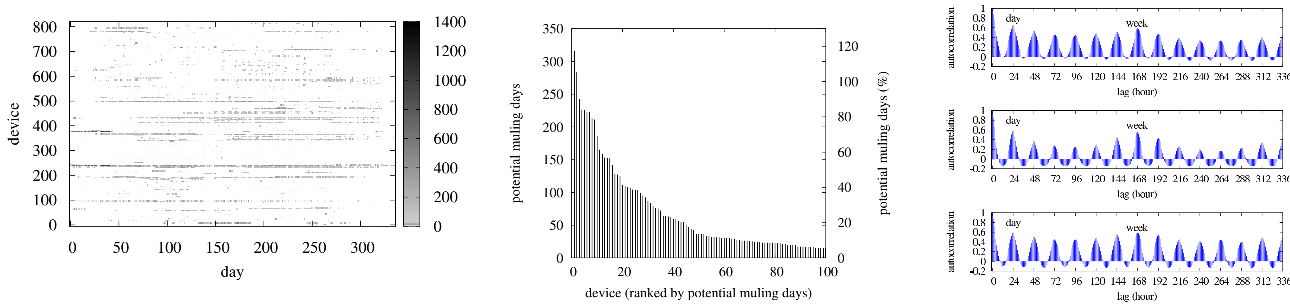
Table 2. Devices seen in observation studies.

Potential sensors: Both traces show that humans naturally encounter many different devices, but most of devices are due to one-time encounters with other mobile devices. (Not surprising since nearly every mobile phone as Bluetooth, while its use outside of phones is growing but lags.) Since we expect most environmental sensors to be stationary, we wish to consider encounters with stationary devices. Each Bluetooth devices includes class information in its public announcements, in addition to its unique MAC address. The class tells what kind of Bluetooth device it is, telephone, headset, and stationary classes such as desktop computer, server computer, modem/gateway, ISDN, loud speaker, set top box, and VCR as stationary. Table 2 identifies these stationary devices and we treat them as *sensor stand-ins*. In addition, 1058 devices omit class information. We find that some devices (both stationary and mobile) send class information occasionally; we classify these as mobile or stationary when possible, or as unknown when they never report.

We conclude that even today, humans encounter many short-range, wireless devices every day. If these devices made sensor data available we would be swimming in data.

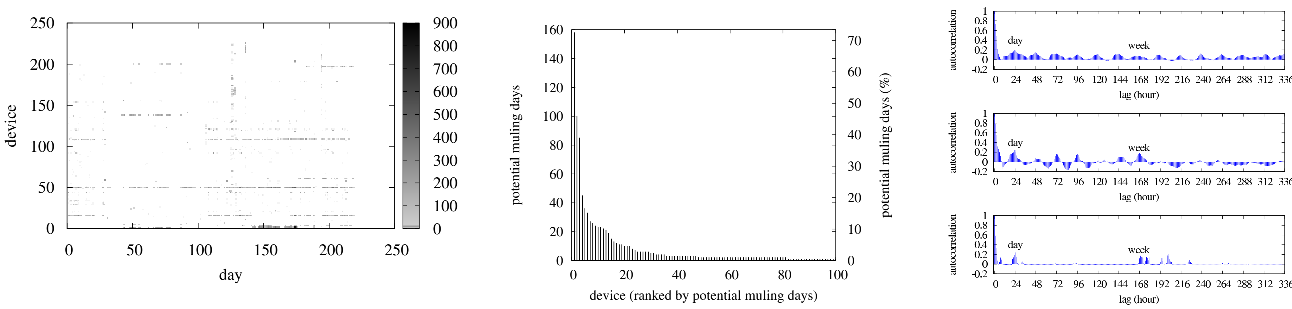
Frequency: Muling depends on frequent encounters with sensors, and humans are only likely to be interested in sensor data they regularly encounter. We therefore next consider how often humans encounter each sensor stand-in.

Figure 2 shows how often each device was seen over the 9 months of Mobility/MIT. Figure 2(a) shows the number of visits for each sensor stand-in (running along the y-axis) on each day (the x-axis). While some sensor stand-ins are seen only a few times, many appear repeatedly. We define a *potential muling day* as a day when some mule sees a given sensor-stand-in at least once. If we assume data can tolerate up to 24 hours of latency, this metric represents muling “timely coverage”. Figure 2(b) shows the number of potential muling days for the top 100 most frequently seen sensor stand-ins. We see that 26 devices have more than 100 potential muling days; these are seen on about 35% or more of the days of the dataset. These devices are excellent candidates for opportunistic muling, without their own wide-area network infrastructure.



(a) Contact frequency per day. (b) Potential muling days by popular sensor. (c) Autocorrelation of contacts between any mule the first, fourth, and sixth most seen sensors.

Figure 2. Contact with sensor stand-ins (dataset: Mobility/MIT).



(a) Contact frequency per day. (b) Potential muling days by popular sensor. (c) Autocorrelation of contacts between any mule the first, fourth, and sixth most seen sensors.

Figure 3. Contact with sensor stand-ins (dataset: Mobility/ISI).

We repeated this study with our second observation dataset (Mobility/ISI). The results in Figure 3 are qualitatively similar to our findings from Mobility/MIT, even through the population of mobile observers was much smaller. Part of the reason for similar coverage with fewer mobile devices is much greater use of Bluetooth today.

We have shown that there are hundreds of sensor stand-ins that are frequently seen with casual movement, and some of these are seen quite frequently. We next look at the regularity of sensor encounters with opportunistic mobility.

5.2 Regularity in Human Mobility

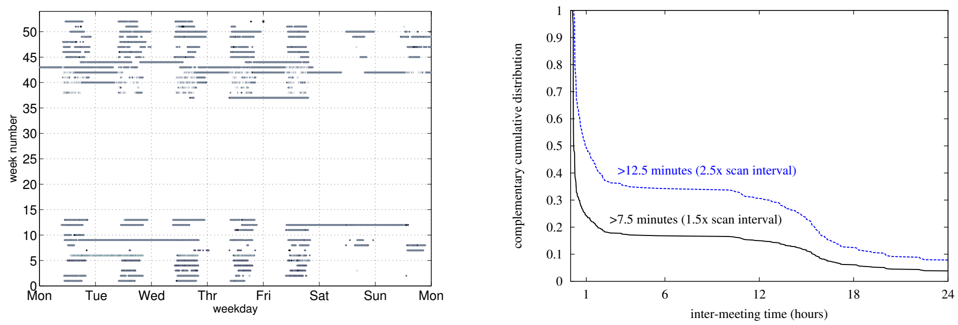
We have shown that there are many potential sensors for which Bluetooth can provide reasonable coverage, and that some are seen many times. But how *regular* is communication? If data muling is to replace wide-area networking, we need guarantees that sensors are seen not only frequently, but regularly. We next re-examine our observation datasets to judge regularity in potential data muling.

To understand how often a specific sensor is visited, Figure 4 shows contact patterns (left) and inter-meeting times (right) for a specific sensor stand-in (sensor-stand-in 270 from Mobility/MIT). We examined other sensors, this kind of contact is typical for sensors that are seen relatively frequently (this sensor was the fourth most common sensor).

Figure 4(a) shows the contact pattern over the dataset. This sensor shows strong regular daily contact on most weekdays (corresponding during a 9am to 5pm workday). It also shows occasional contact on weekends (Saturday and Sunday) and extended periods of contact (for example, at the beginning of the ninth week). We can infer that this target device is located in the person's work area. For other sensors we see patterns of contact during non-work times (9pm through 7am). This data shows the potential for regular contact using mobile-phone-based mules.

To quantify the degree of contact, Figure 4(b) shows inter-meeting times. Since contact is determined only by scans every 5 minutes (for Mobility/MIT), we compute time between contacts by assuming contacts within some window show continuous connectivity. We use two different timeout windows, 1.5× and 2.5× the scan interval (7.5 or 12.5 minutes), to detect gaps, optionally bridging over a single missed scan. Both windows show similar behavior: the most frequent inter-meeting time is 5 or 10 minutes, corresponding to the window size, because it is the minimum timeout we can detect. They also a relatively sharp drop-off around 16-hours, corresponding to the non-work part of a weekday.

To better understand how often sensors are visited, Figures 2(c) and 3(c) show autocorrelations for six different sen-



(a) Contact patterns by week. (b) Inter-meeting time, for two timeout windows.

Figure 4. Contact with a specific sensor stand-in over time (dataset: Mobility/MIT, mule: 35, sensor-stand-in: 270).

sor stand-ins taken from the two datasets. All sensors show very strong daily periodicity (the peak at lag of 24 hours, this peak is second highest to the lag at the scan interval of 5 or 2 minutes). In addition, we often see the next highest peak at one week (168 hours), showing a strong weekly periodicity. In fact, for the sixth most seen sensor stand-in in Mobility/ISI, only immediate, daily and weekly periodicities are strong. This more complete analysis of mobility data confirms regular periodicity in human mobility and strongly suggests the potential for human-based data muling.

6 Evaluation of Our Data Muling System

Having established the potential for data muling, we next explore our data muling implementation. We review how frequent sensor-mule rendezvous are in our four sensornet deployments (§6.1). We then study loiter time to understand how data size and movement interact (§6.2). Finally, we evaluate the energy requirements of muling (§6.4).

6.1 Does Data Muling Work in Real Deployments?

Our analysis of the observation datasets suggest the potential for regular data muling with mobile phones and opportunistic mobility. We next turn to our four deployments: Subsidence/Oilfield, Subsidence/Urban-BT, Subsidence/Urban-Wi-Fi, and People/ISI, to evaluate how muling performs in practice.

6.1.1 Muling in the Oilfield

First we consider the Subsidence/Oilfield deployment. We turned to muling here to monitor experiment hardware, after being denied access to the industrial wireless network and finding the 3G mobile data network unusable. This application requires frequent status updates from the sensors to evaluate hardware operation and trigger rapid maintenance when required. Our muling system used a mobile phone carried by a field engineer who would gather data using the phone's Bluetooth connection, then carry the phone into town at night where 3G coverage is quite good. The alternative to muling was a 6-hour round-trip drive into the field at periodic intervals; our goal instead was at least weekly updates and data-to-date.

This experimental deployment brings three conclusions: it requires intentional movement, muling meets our latency expectations, and, in §6.1.2 it succeeds in reducing downtime. We consider each next.

First, this deployment requires *intentional*, not opportunistic movement. Our studies of the observational datasets show that many sensor stand-ins are seen often. While true, in the Subsidence/Oilfield deployment we have only *one* mule and four *specific* sensors to visit, not 3 or 100 mules scanning for any sensors. The oilfield is a very large area (more than 50 km^2), and Bluetooth radios typically operate with range limits of 10 m or less, so our single mule would nearly *never* meet our sensors accidentally. Instead, we asked the field engineer to *intentionally* travel to the sensors and wait next to them while data transfer takes place. As Figure 5(a) shows, regular weekly muling with intentional mobility was successful for the first 5 weeks of operation. However, the figure also shows a weakness in relying on human mobility: the gap in weeks 8 through 10 is due to December vacation by our mule carrier.

In a large outdoor area, muling with opportunistic mobility requires many more mules and longer-range radios. The site has many workers moving about, if we could equip each engineer's phone with muling, or instrument company trucks with Wi-Fi-based mules, we expect we would reduce the need for intentional mobility.

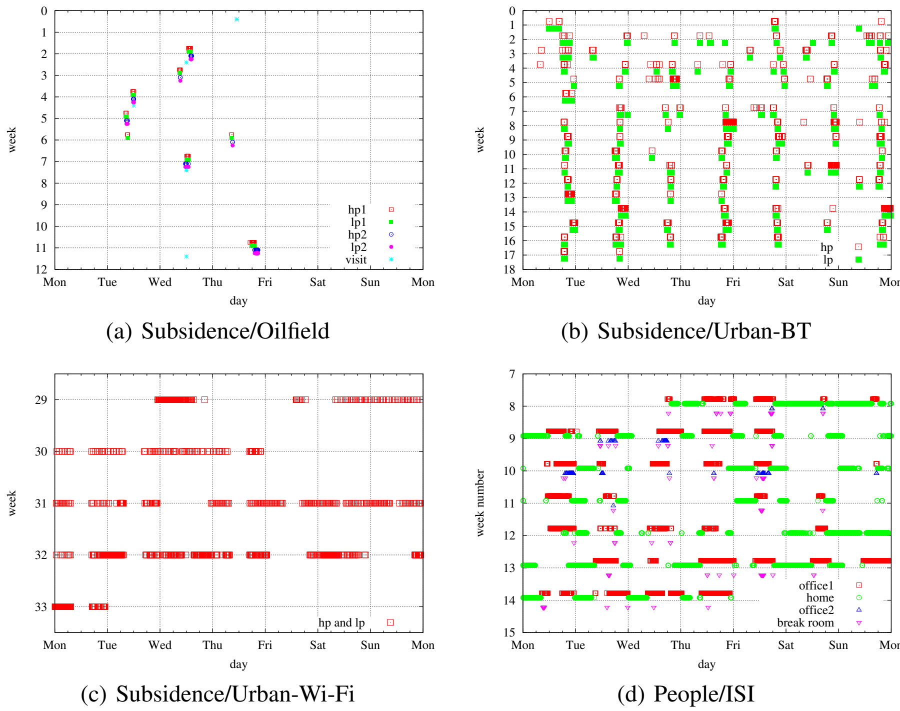
Second, muling meets our latency expectations, as Figure 5(a) shows, regular weekly muling with intentional mobility was successful. While expected (due to intentional mobility), this result confirms that our muling system meets its goal, and that our system is usable by a non-expert.

6.1.2 Oilfield Muling and Sensor Coverage

We have shown muling works in the oilfield, but does it help? Muling's benefits are providing data and informing us of deployment problems more rapidly than periodic visits.

The alternatives are muling every week, or memory card swaps every three weeks. On average, muling retrieves data 1.5 weeks before it would have been acquired with scheduled visits. Although both muling and memory card swaps require a visit to the site, wireless muling is much simpler. Card exchanges require a level of technical involvement that field personnel were unwilling to undertake.

More important than getting data more rapidly, muling alerted us to problems with our deployment. As an experimental deployment, we encountered several hardware and



(a) Subsidence/Oilfield (b) Subsidence/Urban-BT

(c) Subsidence/Urban-Wi-Fi (d) People/ISI

Figure 5. Weekly patterns of four muling deployments. Each symbol is a muling opportunity for a specific sensor.

sensor	outage dates	detection date mule	detection date scheduled	dur. unknown mule	dur. unknown scheduled
hp1	Nov. 12 to Nov. 23	Nov. 17	Nov. 23	5	11
	Nov. 26 to Dec. 15	Nov. 15	Dec. 15	4	19
hp2	Nov. 12 to Nov. 23	Nov. 17	Nov. 23	5	11
	Dec. 09 to Dec. 15	Dec. 15	Dec. 15	6	6
lp1	none	—	—	0	0
lp2	Nov. 10 to Nov. 23	Nov. 17	Nov. 23	7	13
total				**27**	**60**

Table 3. Outages in Subsidence/Oilfield.

software problems. The early detection of muling gave us an opportunity to understand when outages occur. Table 3 shows outages dates in our deployment, and when these outages were detected with muling, compared to when they would have been detected with a scheduled visit. As can be seen, muling halved the time outages were unknown, from 60 sensor-days to 27, allowing us to make an informed decision about the need for early field visits for maintenance.

6.1.3 *Muling for Urban Sensing*

Bluetooth: Our second deployment is Subsidence/Urban-BT deployment. While the application is the same as Subsidence/Oilfield, this site is an urban area, and the goal is a controlled experiment rather than field data.

This experiment again required intentional mobility. Although located at a residence, the sensors are behind the garage and so usually of Bluetooth range from the primary living areas. As Figure 5(b) shows, intentional mobility addresses this problem. However, in examining the data we note about 7% of the contacts appear to be opportunistic. Opportunistic mobility is therefore possible and may provide a benefit even when not planned. We expect use of Wi-Fi over Bluetooth would have allowed all-opportunistic muling in this scenario.

Wi-Fi: In the *Subsidence/Urban-Wi-Fi* deployment replace Bluetooth with 802.11. The longer radio range of 802.11 makes data muling possible with all-opportunistic mobility.

Figure 5(c) shows that data mule covers sensors most of the time during four weeks of experiment period. Sensors are contacted with data mule opportunistically, because we do not employ intentional mobility used in the Subsidence/Urban-BT deployment. We conclude pure opportunistic mobility is sufficient for data muling in the Subsidence/Urban-Wi-Fi deployment. In addition, we analyze how a faster data rate of 802.11 affects in muling data in §6.3. (Both sensors in Figure 5(c) have the same contact pattern, because both sensors share a common access point.)

6.1.4 *Muling for Office Sensing*

Our final deployment was designed to test purely opportunistic data muling, and with small data sizes. We placed several sensors in an office environment better suited to Bluetooth's short range. Here the sensors sense people (in our implementation, by looking for human-carried Bluetooth devices using a second Bluetooth adapter); we place sensors in four locations: two offices, each visited by a few people; a break room visited by many people. We also place one sensor at an apartment shared by several people. One individual carried the muling device, visiting the apartment and one office and the break room daily, the other office weekly.

Figure 5(d) shows muling opportunities at each sensor over the course of eight weeks. We see many opportunities to mule at office1 and home (square and circle), and regular opportunities at the break room and office2. There is actually some correlation between breakroom and office2 because they are at the edge of Bluetooth range. We conclude that opportunistic muling works very well when radio range and mobility patterns are well matched, as in an office environment.

6.2 Loiter Time Effects on Muling

For data muling to be successful, the mule must stay within radio range of the sensors long enough to transfer any pending data: the loiter time must be longer than the muling time. Muling time is a function of the size of each data item and the number queued up to send, which in turn depends on contact frequency. We next evaluate muling time and estimate loiter times to see how often successful muling is likely to occur.

Observing: We first estimate required muling time as a function of data size and number of queued data items in Figure 6(a). In this graph, each diagonal line represents a single data size, from 1 byte to 1 MB, and each point on that line a different number of data items, from 1 to 100 (for small sizes), or to 20 or 2 for the largest sizes. (Note that the observations in our subsidence applications are 1–2 MB in size, the largest size listed.) Each point is taken experimentally and is the mean of 10 measurements, with error bars showing standard deviation. We include all delays in muling: There is roughly fixed-duration overhead for a mule to discover sensors, and setup a Bluetooth connection, and determine if there is data to mule; together this setup requires about 17 seconds. Then the time to transfer data items is roughly linear with the quantity of data transferred.

Modeling: To understand muling time across data sizes for many possible applications, we fit a simple linear model to our observations. Muling time consists of three components: communication overhead (discovery and connection), data transfer time, and muling overhead. Communication overhead is almost constant; we measure 13 s for Bluetooth discovery and 1 s to open a connection, consistent with what is stated in the Bluetooth specification [2]. Transfer time changes according to the total size of data, and the muling overhead increases according to the number of files. If we define k_{xfr} is the transfer rate, k_F as the per-file overhead, and k_D as monitoring and disconnect overhead, we can solve for these constants using multiple linear regression.

$$\begin{aligned} T_{muling} &= t_{discovery} + t_{conn} + t_{transfer} + t_{muling\_overhead} \\ t_{transfer} + t_{muling\_overhead} &= (1/k_{xfr})SN + k_F N + k_D \\ &\text{where } S \text{ is the data item size} \\ &\text{and } N \text{ is the number of files} \end{aligned} \tag{1}$$

Predicting: This analytic model helps us evaluate how muling can work with different kinds of mobility patterns. The large rendezvous time is a critical factor to muling: with Bluetooth, muling *any* amount of data takes at least 20 s. This limit places a bound on user movement: with a 10 m radio range, a user can move at most 1 m/s if a scan begins immediately on entering radio range. Typical human walking pace depends on age, but ranges from 1.35 to 1.5 m/s for an already walking individual, depending on age [6]. Therefore Bluetooth-based muling will *not work well* for constantly moving humans, even if mules constantly scan for sensors. This result suggests that muling must involve either accidental loitering, longer range radios (perhaps higher power Bluetooth with ranges to 100 m, or 802.11 with ranges of 35–70 m), or great improvements to device discovery protocols. This result is consistent with our observations in People/ISI, since we place each sensor at a location where the data mule is likely to loiter—offices, a break room, or home.

variable	Bluetooth	802.11
$t_{discovery}$	$\mu = 13.01\,\text{s}, \sigma = 0.448\,\text{s}$	$\mu = 3.70\,\text{s}, \sigma = 0.192\,\text{s}$
t_{conn}	$\mu = 1.616\,\text{s}, \sigma = 0.661\,\text{s}$	$\mu = 0.277\,\text{s}, \sigma = 0.179\,\text{s}$
k_{xfr}	1.143 Mb/s	14.435 Mb/s
k_F	0.3280 s/file	0.3642 s/file
k_D	3.402 s	1.670 s

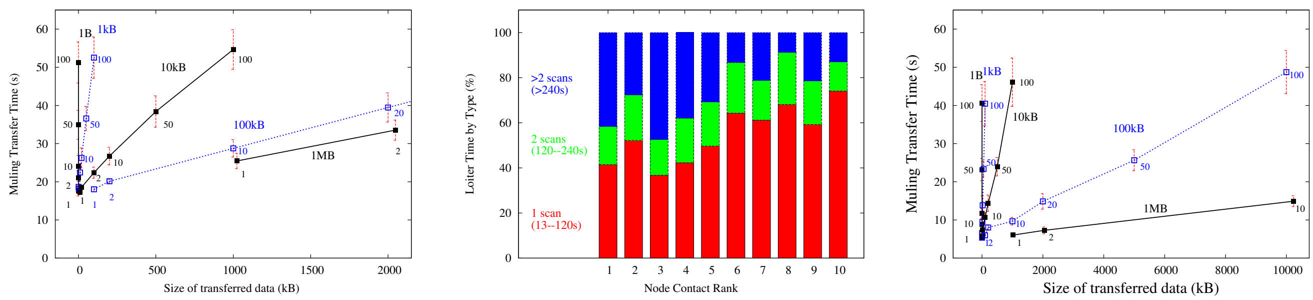
Table 4. Parameters for the data muling time model (802.11 model is discussed in §6.3) .

The model also shows that transfer time is irrelevant for almost all small data sizes (data items than 100 kB), since connect time dominates. However, with the large, 1–2 MB data items in our Subsidence applications, transfer time is very noticeable and transfer time dominates muling time when more than a few of items are queued for transfer.

Finally, we can compare this model to our observation datasets to evaluate how often muling would likely succeed. This comparison is difficult, because those datasets record only sensor-stand-in *detection* times, not contact times, so we know when a device was seen but not for how long. However, if we assume multiple consecutive detections correspond to continuous contact, then we can infer contact times in those cases. If consider the Mobility/ISI dataset since it has greater scan frequency than Mobility/MIT (2 rather than 5 minutes between scans), we can then classify more than two scans (loiter time more than 4 minutes) as enough time to mule almost anything, two scans (2 to 4 minutes) as enough time to mule all cases on Figure 6(a), and one scan (up to 2 minutes) as possibly enough to mule something

Figure 6(b) shows loiter times in these three categories for the ten most frequently contacted sensor stand-ins from Mobility/ISI. For these sensors, about 20%–60% of contacts are 2 or more than 2 scans; long enough to transfer at least 15 MB of data. We cannot judge loiter times for single scans, but there still seems some chance to transfer smaller data items (1–100 kB) in 17–20 s.

Large, multi-item transfers: Finally, our focus here has been understanding how brief loiter times interact with minimal, opportunistic data transfer. In our Subsidence applications, data items are each 1–2 MB in size, and muling is done once a day (Subsidence/Urban) or once a week (Subsidence/Oilfield). Each sensor generates an observation every 2 hours, so these applications require moving 12–24 MB or 84–168 MB of data per muling session. Our model predicts muling times of 2–20 minutes, and our experiences bear this out: muling takes a long time for large-data sensing with infrequent rendezvous. We therefore replace Bluetooth with Wi-Fi in Subsidence/Urban-Wi-Fi as described below (§6.3).



(a) The transfer time between mule and sensor with different data sizes via Bluetooth (b) Loiter time of mules with 10 most frequently contacted nodes (c) The transfer time between mule and sensor with different data sizes via Wi-Fi

Figure 6. Data muling time and loiter time (dataset: Mobility/ISI)

Small data sizes: Finally, we observe that many sensornet applications, like the People/ISI, have data sizes that are tens of bytes instead of megabytes. As shown in Figure 6(a), small data items can be transferred so quickly that dozens can easily be transferred with even a brief rendezvous. Our current software has a large per-item relay overhead; batching could improve performance further.

6.3 Faster Data Communication with 802.11

In the previous section we evaluated the muling time with Bluetooth. Although Bluetooth works, the data transfer rate is slow, with some scenarios requiring loiter times of 70–80 minutes. The short radio range of Bluetooth is also problematic in the Subsidence/Urban-BT deployment where it forces intentional mobility for mule-sensor rendezvous.

Here we explore 802.11 as an alternative to Bluetooth for data muling communication, replacing Bluetooth with Wi-Fi in our Subsidence/Urban-Wi-Fi deployment. We see that the use of 802.11 permits shorter loiter times and provides longer communications range, allowing Subsidence/Urban-Wi-Fi to work with only opportunistic muling; intentional mobility is no longer required.

First, 802.11 has much faster data rate and shorter discovery time than Bluetooth. These improvements significantly reduce the required loiter time to get data from sensor. We conduct the same analysis with 802.11 (shown in Figure 6(c)) and fit it to our analytic model of muling time (shown in Equation 1) to evaluate the muling performance improved by the faster data rate and shorter discovery time.

Table 6.2 compares computed model parameters for muling time for both Bluetooth and 802.11. With 802.11, discovery time is reduced to 3.7 s, less than one-third of the time with Bluetooth discovery (13.0 s). It also takes only 0.3 s to connect to a sensor via 802.11, where Bluetooth takes 1.6 s. These short discovery and connection times allow data mules not only scan neighboring sensors rapidly, but also check the availability of new data quickly.

The transfer rate with 802.11 is 12 times faster than Bluetooth (shown in Table 6.2). This faster data rate enables data muling during rendezvous periods that would be too short for Bluetooth. For example, a Bluetooth mule takes 23 s to mule 1MB of data, so with Bluetooth, loiter time must be nearly half a minute. However, more than half of contacts are only 13–120 s (in Figure 6(b)). In those short contacts, Bluetooth may discover the sensor (longer than 13 s), but then be unable to complete the data transfer (when loiter time is less than 23 s). With 802.11, a mule requires only 6 s to rendezvous and transfer 1 MB of data. Thus, data muling with 802.11 is quick enough to complete the data transferring process, even for short contacts.

The improvement is even larger when muling large and multiple data items. In the Subsidence/Oilfield deployment, it takes 20 minutes to transfer a week's worth of data from a sensor with Bluetooth (84–168MB). A field engineer who carries a data mule with Bluetooth will have to spend 70–80 minutes to mule data from all four sensors. With 802.11 muling, data transferring time will take 1–2 minutes per sensor and less than 6 minutes for all, so the field engineer no longer need to wait hours to mule large and multiple data items, addressing a concern raised in Subsidence/Oilfield.

As importantly, the long ranges of 802.11 (35–70 m) allow more frequent opportunistic muling. In the Subsidence/Urban-BT deployment, sensors are located 10–15 m away from the typical rooms where the data mule is kept. Since the Bluetooth radio's range is only 10 m, opportunistic muling is quite rare—only 7% of all contacts (§6.1.3). When we replace Bluetooth with 802.11 (in the Subsidence/Urban-Wi-Fi deployment), *all* 2638 contacts for the four-week deployment were opportunistic.

Together, shorter loiter times and opportunistic mobility greatly improve the easy-of-use for muling.

6.4 Energy Consumption

Energy consumption of mobile devices can be critical for use. Although many mobile phones are charged every day, running out of power mid-day quickly draws user complaints. We next evaluate the energy consumption of our data muling system.

To measure energy consumption, we observe energy consumption over one hour of operation. Over that time we carry out regular scans for sensors every two minutes, and we relay 12 datafiles, each 1 MB in size. This amount of data is roughly equivalent to the amount of data sent in one daily rendezvous for the Subsidence/Urban-BT deployment. While doing muling we run the PowerTutor application [31] to record total energy consumption, and at the end of experiment we use Android's Settings: About Phone: Battery Use application to evaluate the percentage energy consumed by

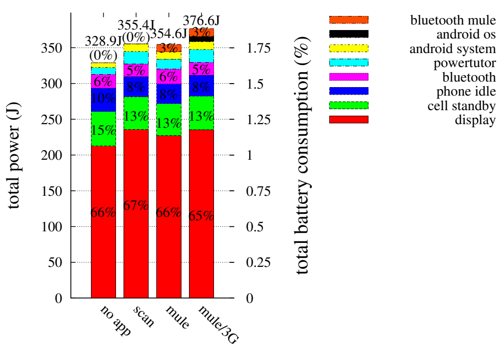


Figure 7. Energy consumption of the Data Mule application over one hour.

each component. The data we report is the result of one set of experiments on a Samsung Galaxy S phone running Android 2.1 and BluetoothMule 2.2.3. We see similar proportions of energy use on a HTC Magic (branded as T-Mobile myTouch 3G) running Android 2.2.1. Finally, we check phone status for the first and last five minutes of the period, activating the display.

Figure 7 shows both total energy consumption (the heights of the bars, and indicated numerically above each bar and on the left scale), component costs (shown as percentages for key components in each bar), and relative consumption of one hour of operation run compared to total battery (the right scale). We consider four scenarios: the standard phone without muling ("no app"), running muling doing scanning with one sensor in range ("scan"), muling the amount of data ("mule"), and muling the data and sending it over the 3G network to the Internet ("mule/3G").

Our first observation from this analysis was surprising to us: the display is by *far* the largest energy consumer. An early version of our mule intentionally left the display active to inform the user of progress; we quickly removed this energy-wasting choice.

Second, we see that frequent scanning consumes a noticeable amount of energy: about 30 J, comparing the scan vs. no-app bars. Over the course of 24 hours, scanning consumes about 5% of total battery energy.

We see that scanning takes noticeable energy. Whether or not energy consumption from scanning is a problem depends on if it runs the phone out of battery before the phone is recharged. In our use we found that scanning did not frequently exhaust phones batteries, however, on occasions when the phone was taxed for other reasons (say, long voice calls), muling contributed to forcing an early recharge.

By contrast, muling the data from the sensor does *not* consume much additional energy (compare the scan and mule bars in Figure 7). Bluetooth is optimized for energy-efficient short-range data movement, while the cost of listening many seconds to scan for potential devices is much more expensive. (This trade-off is the same one that prompted low-power listening [15, 26] and scheduling [37, 33, 38] optimizations in MAC protocols.)

Finally, we see that 3G wide-area communication approximately doubles energy consumption, consuming another 22 J (compare mule/3G vs. mule). In fact, this experiment was conducted in an urban area, and the energy costs of 3G can be greater still in areas where cellular coverage is poorer.

We conclude that energy costs of muling are noticeable, however they are relatively small in absolute terms compared to the primary function of most mobile phones: taking calls and communicating information, as reflected in the cellular standby and display energy costs. Energy consumption of muling must be considered, but in our experience it is usually not a primary factor in phone usability. In addition, when muling is the primary use for the phone, as it was in our Subsidence/Oilfield deployment, energy consumption is well within the capacity of today's typical phones if they recharge every day.

Based on our experience running muling we implemented several features to manage power usage. First, our muling program disables relay to the Internet when operating on battery is less than 20% (however, muling from sensor-to-phone is still done). This addition implements the policy of "phone first, Internet second", on the assumption that the phone will likely be connected to grid power shortly and can complete data relay to the Internet at that time.

7 Related Work

Our work builds on prior work in data sharing applications, data muling in sensor networks, and understanding human mobility.

7.1 Data Muling in Sensor Networks

The concept of using data mules to support sparse sensor networks is an old one [21, 9, 28, 25, 4, 29, 20, 35]. The key idea is that a mobile mule can provide energy efficient data relay with a short range radio, while mobility can bridge long distances. Muling schemes can be categorized by the type of mobility they expect: *random*, such as with animals [21], humans [18], or simulated [28]; *controlled*, with robots [29], airplanes [35], or boat; an *predictable* mobility with trains, shuttles, or buses [9, 25], or semi-predictable farmworkers [4].

Different expectations about mobility usually are reflected by different research methodologies. Some prior studies of humans as data mules have modeled primarily random walk or random waypoint mobility patterns [28], yet we know human mobility is far from random. Other work has considered semi-random models [21, 29, 35]. We instead study human mobility using traces from mobile phones, sampling the mobility of real humans. We also and evaluate the feasibility data muling with real human movement in our four mule-based sensornet deployments.

Closest to our work, Burrell et al. [4] and CenWits [18] recognize the potential of human mobility in data muling from sensors We compares these work with ours next.

Burrell et al. study use of sensor networks in vineyards [4]. Based on ethnographic studies, they identify farmworker mobility as capable of supporting data muling, similar to our identification of field engineer movement. Their work focuses on motes dedicated for muling, so they do not explore mobile phone traces, and their field system required sensor densities that eliminated the need for data muling [1].

CenWits is a search-and-rescue system for hikers using hiker-carried motes [18]. They recognized the need for communication in sparse areas, but as with Burrell, they propose a dedicated system.

Several recent applications have explored participatory sensing using mobile-phones [3]. These applications applications often focus on the mobile phone as the sensor, while the applications we identify in §2.2 assume mobile phones relay data from an in-situ sensor. GarbageWatch is one proposed application: participants take photos of garbage bins on a campus to improve recycling [32]. We instead focus on data muling to permit reports on remote garbage bins.

7.2 Understanding Human Mobility

Many groups have studied human mobility for data communications, we summarize three very relevant studies here. As described above, Burrell et al. explored farmworker mobility with ethnographic studies [4].

Chaintreau et al. study transfer opportunities using wireless devices carried by humans as we do [8]. They find heavy-tailed inter-meeting times, and so recommend new opportunistic forwarding algorithms between mobile nodes. We discuss opportunistic data forwarding as possible future work for our system (§3.4), and instead focus on the data transfer between mobile phones and stationary nodes with a real working system.

Eagle et al. study human mobility patterns using mobile phones as proxies [14]. Their goal is to understand human and group relationships. They find that individual behavior over a specific day can be approximated by a weighted sum of repeated behavioral patterns they call eigenbehaviors, corresponding to behaviors such as a normal day vs. traveling, weekends vs. weekday, etc. We build on their studies of human mobility to evaluate the feasibility of data muling in daily life. While they focus on extracting and analyzing the underlying structure of human behaviors, we exploit the routine of human mobility to cover sparsely deployed sensors efficiently.

8 Future Work

Although our system has been operational for some time, there remain several areas of future work.

As proposed in CarTel [19], a system should make the most of short periods of opportunistic muling. We currently implement a simple policy of relaying newest-data first, but summaries or sampling may be better policies for specific applications.

We have several ideas to make muling rendezvous more efficient. First, there are opportunities to improve protocol-level rendezvous, using techniques such as low-power listening [15, 26] or scheduling [37, 33, 38]. Also, we believe that study of prior mobility patterns can improve predictions about future rendezvous, allowing us to dynamically alter protocol-level behavior when conditions for rendezvous are favorable. In addition to speeding rendezvous, mules that visit sensors very frequently waste energy confirming there is nothing new to relay. We therefore may explore less frequent scanning for often-visited sensors.

Our system focuses on human mobility, so human motivation plays a role. While in some cases (like Subsidence/Oilfield), muling may be company policy, in more general cases we need to consider reasons for users participate in muling. More importantly, our current system makes no attempt to influence human movement. We would like to explore giving users incentives to approach sensors to assist muling.

Finally, while we focus on discrete sensor data, with data item sizes of bytes to megabytes, generated at fixed intervals, still or video cameras represent another class of sensors is high-data rates and non-fixed intervals. With Wi-Fi, muling may well suited to retrieve large data streams from remote areas.

9 Conclusion

We have shown that data muling with human-carried mobile phones is both possible and practical. We have demonstrated the potential with analysis of two datasets of mobile phone movement, showing that individuals see many potential sensors, and see some regularly. Inspired by this potential, we implemented a data muling system and used it to share data in four deployed sensornets. We showed that short radio ranges of Bluetooth require intentional mobility to make muling practical for industrial and even some urban applications, but that opportunistic muling is suitable for our office-based deployment. We investigated trade-offs in data size, visitation frequency, and how they interact with muling and loiter times, and we examined energy consumption. While work remains, we believe data muling has a role in bringing communication to sparsely connected sensors.

Acknowledgments

This work would not have been possible without the strong support of several Chevron personnel in cooperation with CiSoft. We thank Greg LaFramboise for project guidance and Tai Nguyen for support at Lost Hills. We also thank John Houghton and Roy Kutlik for involvement with Subsidence/Oilfield, and Iraj Ershagi (USC) and Mike Hauser for their guidance as CiSoft directors. Finally, at USC/ISI, Fabio Silva lead Subsidence/Oilfield development, and Zhihan (Sam) Lu and Hsing-hua Chen did early development and experiments; we thank them for their essential role in the subsidence deployments.

10 References

[1] R. Beckwith. Personal communication, Apr. 2011.

[2] Bluetooth Special Interest Group. Bluetooth specication version 1.1 and 1.2, 2001.

[3] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava. Participatory sensing. In *Proc. of the 1st ACM World Sensor Web Workshop*, pages 117–134, Boulder, Colorado, USA, Nov. 2006. ACM.

[4] J. Burrell, T. Brooke, and R. Beckwith. Vineyard computing: Sensor networks in agricultural production. *IEEE Pervasive Computing Magazine*, 3(1):38–45, Jan. 2004.

[5] V. Bychkovsky, B. Hull, A. K. Miu, H. Balakrishnan, and S. Madden. A measurement study of vehicular internet access using in situ wi-fi networks. In *12th ACM MOBICOM Conf.*, pages 50–61, Los Angeles, CA, September 2006.

[6] N. Carey. Establishing pedestrian walking speeds (draft results). Technical report, Portland State University, 2005.

http://www.westernite.org/datacollectionfund/2005/psu_ped_summary.pdf.

[7] A. Cerpa, J. Elson, D. Estrin, L. Girod, M. Hamilton, and J. Zhao. Habitat monitoring: Application driver for wireless communications technology. In *2001 ACM SIGCOMM Workshop on Data Communications in Latin America and the Caribbean*, pages 20–41, San Jose, Costa Rica, Apr. 2001.

[8] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott. Impact of human mobility on the design of opportunistic forwarding algorithms. In *Proc. IEEE Infocom*, 2006.

[9] A. Chakrabarti, A. Sabharwal, and B. Aazhang. Using predictable observer mobility for power efficient design of sensor networks. In *Proc. of the 2nd Intl. Workshop on Information Processing in Sensor Networks (IPSN)*, pages 129–145, 2003.

[10] K. Chang, N. Yau, M. Hansen, and D. Estrin. SensorBase.org—a centralized repository to slog sensor network data. In *Proc. of the IEEE Euro-American Workshop on Middleware for Sensor Networks*. IEEE, June 2006.

[11] B. Chen, K. Jamieson, H. Balakrishnan, and R. Morris. Span: an energy-efficient coordination algorithm for topology maintenance in ad hoc wireless networks. In *Proc. of the ACM International Conference on Mobile Computing and Networking*, pages 85–96, Rome, Italy, July 2001. ACM.

[12] X. Chen. Enabling disconnected transitive communication in mobile ad hoc networks. In *Proc. of the Workshop on Principles of Mobile Computing*, pages 21–27, 2001.

[13] B. Collins. The true cost of publishing on the Amazon Kindle. http://www.pcpro.co.uk/blogs/2011/02/15/the-true-cost--of-publishing-on-the-amazon-kindle/.

[14] N. Eagle and A. S. Pentland. Eigenbehaviors: identifying structure in routine. 63(7):1057–1066, May 2009.

[15] A. El-Hoiydi, J.-D. Decotignie, and J. Hernandez. Low power MAC protocols for infrastructure wireless sensor networks. In *Proc. of the Fifth European Wireless Conference*, pages 563–569, Barcelona, Spain, Feb. 2004. Feb.

[16] Google. Google PowerMeter. http://www.google.org/powermeter/, 2007.

[17] D. Hammerstrom. Olympic peninsula gridwise study, 2007.

[18] J.-H. Huang, S. Amjad, and S. Mishra. CenWits: a sensor-based loosely coupled search and rescue system using witnesses. In *Proc. of the Third ACM SenSys Conference*, pages 180–191, San Diego, California, USA, Nov. 2005. ACM.

[19] B. Hull, V. Bychkovsky, Y. Zhang, K. Chen, M. Goraczko, A. Miu, E. Shih, and S. M. Hari Balakrishnan. CarTel: A distributed mobile sensor computing system. In *Proc. of the Fourth ACM SenSys Conference*, pages 125–138, Boulder, Colorado, USA, Nov. 2006. ACM.

[20] S. Jain, R. C. Shah, W. Brunette, G. Borriello, and S. Roy. Exploiting mobility for energy efficient data collection in wireless sensor networks. *Mob. Netw. Appl.*, 11(3):327–339, 2006.

[21] P. Juang, H. Oki, Y. Wang, M. Martonosi, L. shiuan Peh, and D. Rubenstein. Energy-efficient computing for wildlife tracking: design tradeoffs and early experiences with ZebraNet. In *Proc. of the 10th international conference on Architectural support for programming languages and operating systems*, ASPLOS-X, pages 96–107, 2002.

[22] R. Katz, D. Culler, S. Sanders, S. Alspaugh, Y. Chen, S. Dawson-Haggerty, P. Dutta, M. He, X. Jiang, L. Keys, A. Krioukov, K. Lutz, J. Ortiz, P. Mohan, E. Reutzel, J. Taneja, J. Hsu, and S. Shankar. An information-centric energy infrastructure: The Berkeley view. *Sustainable Computing: Informatics and Systems*, 2011.

[23] J. Lebrun, C.-N. Chuah, D. Ghosal, and M. Zhang. Knowledge-based opportunistic forwarding in vehicular wireless ad hoc networks. In *61st IEEE Vehicular Technology Conference*, volume 4, pages 2289–2293, 2005.

[24] Microsoft. MS Hohm. http://www.microsoft-hohm.com/, 2009.

[25] A. S. Pentland, R. Fletcher, and A. Hasson. DakNet: Rethinking connectivity in developing nations. *IEEE Computer*, 37(1):78–83, Jan. 2004.

[26] J. Polastre, J. Hill, and D. Culler. Versatile low power media access for wireless sensor networks. In *Proc. of the Second ACM SenSys Conference*, pages 95–107, Baltimore, MD, USA, Nov. 2004. ACM.

[27] C. Schurgers, V. Tsiatsis, S. Ganeriwal, and M. Srivastava. Optimizing sensor networks in the energy–latency–density design space. *IEEE Transactions on Mobile Computing*, 1(1):70–80, Jan. 2002.

[28] R. C. Shah and S. Roy. Data mules: Modeling a three-tier architecture for sparse sensor networks. In *Proc. of the IEEE SNPA Workshop*, pages 30–41, 2003.

[29] A. A. Somasundara, A. Kansal, D. D. Jea, D. Estrin, and M. B. Srivastava. Controllably mobile infrastructure for low energy embedded networks. *IEEE Transactions on Mobile Computing*, 5(8):958–973, 2006.

[30] R. Szewczyk, A. Mainwaring, J. Polastre, J. Anderson, and D. Culler. An analysis of a large scale habitat monitoring application. In *Proc. of the 4th ACM SenSys Conference*, pages 214–226, Baltimore, Maryland, USA, Nov. 2004. ACM.

[31] L. Z. B. Tiwana, Z. Qian, Z. Wang, R. P. Dick, Z. M. Mao, and L. Yang. Accurate online power estimation and automatic battery behavior based power model generation for smartphones. In *Proc. of the 8th IEEE/ACM/IFIP Intl. Conference on Hardware/Software Codesign and System Synthesis*, pages 105–114, New York, NY, USA, 2010. ACM.

[32] UCLA CENS. Garbagewatch. http://urban.cens.ucla.edu/projects/gabagewatch/.

[33] T. van Dam and K. Langendoen. An adaptive energy-efficient MAC protocol for wireless sensor networks. In *Proc. of the First ACM SenSys Conference*, pages 171–180, Los Angeles, California, USA, Nov. 2003. ACM.

[34] H. Wang, D. Estrin, and L. Girod. Preprocessing in a tiered sensor network for habitat monitoring. In *EURASIP JASP special issue of sensor networks*, pages 392–401, 2002.

[35] Y. Wang and C. hua Wu. Robot-assisted sensor network deployment and data collection. In *Proc. of the Computational Intelligence in Robotics*, pages 467–472, 2007.

[36] Y. Xu, J. Heidemann, and D. Estrin. Geography-informed energy conservation for ad hoc routing. In *Proc. of the ACM International Conference on Mobile Computing and Networking*, pages 70–84, Rome, Italy, July 2001. ACM.

[37] W. Ye, J. Heidemann, and D. Estrin. Medium access control with coordinated, adaptive sleeping for wireless sensor networks. *ACM/IEEE Transactions on Networking*, 12(3):493–506, June 2004.

[38] W. Ye, F. Silva, and J. Heidemann. Ultra-low duty cycle MAC with scheduled channel polling. In *Proc. of the Fourth ACM SenSys Conference*, pages 321–333, Boulder, Colorado, USA, Nov. 2006. ACM.