Leveraging Social System Networks in Ubiquitous High-Data-Rate Health Systems

Tammara Massey, *Member, IEEE*, Gustavo Marfia, Adam Stoelting, Riccardo Tomasi, Maurizio A. Spirito, *Member, IEEE*, Majid Sarrafzadeh, *Fellow, IEEE*, and Giovanni Pau

***Abstract*—Social system networks with high data rates and limited storage will discard data if the system cannot connect and upload the data to a central server. We address the challenge of limited storage capacity in mobile health systems during network partitions with a heuristic that achieves efficiency in storage capacity by modifying the granularity of the medical data during long intercontact periods. Patterns in the connectivity, reception rate, distance, and location are extracted from the social system network and leveraged in the global algorithm and online heuristic. In the global algorithm, the stochastic nature of the data is modeled with maximum likelihood estimation based on the distribution of the reception rates. In the online heuristic, the correlation between system position and the reception rate is combined with patterns in human mobility to estimate the intracontact and intercontact time. The online heuristic performs well with a low data loss of 2.1%–6.1%.**

***Index Terms*—Biomedical informatics, body sensor networks (BSNs), mobile ad hoc networks, mobile health systems, preventive interventions, public health informatics, social system networks, wireless ad hoc networks.**

I. Introduction

MOBILE body sensor networks (BSNs) consist of lightweight embedded systems and medical sensors that aid in the treatment and monitoring of many diseases and disorders, such as neurological disorders and chronic cardiac diseases. Patients with mobile lifestyles can benefit from continuous monitoring, early detection, and prevention of complications. Early detection of complications result in better recovery rates and lower overall health care costs.

Social system networks research aims to understand the interconnections, structure, dependencies, and patterns in human mobility using embedded systems. These mobile social networks raise a challenge in communication due to network partitions. During network partitions, these lightweight systems with limited storage capacity cannot upload data to the central server.

Our goal is to use network-analysis techniques to model complex social health care networks and to efficiently enable continuous data collection on lightweight embedded systems. Our research has two main contributions. First, a new concept, social system networks, is presented. *Social system networks* are mobile networks of lightweight systems that extract patterns in interdependent social relationships to improve system design. Patterns in social networks are extracted and used to estimate intercontact time and intracontact time between patients and base stations. Second, a model of the reception rate is leveraged in the global algorithm for maximum likelihood estimation (MLE) and in the local heuristic for the online detection of the system's position on the patient. This local heuristic allows a continuous stream of data to be collected and reduces the amount of data lost on the system.

The intended model applies to the continuous monitoring of patients with high-data-rate sensors in a nursing home. The number of patients can range from tens to hundreds of patients spread out over several nursing home facilities. A high-data-rate electrocardiogram (ECG) sensor data collects data using a social system network. This system is specifically well suited for elderly or chronic disease patients that constantly need to be monitored for lengthy period of time ranging from weeks to months. However, these active patients still have normal, mobile lifestyles and cannot be tethered to a large system.

All mobile networks (cellular, 802.11, Bluetooth, Zigbee, etc) suffer from network partitions. It is unrealistic to assume that patients will always be in an urban area with well-developed resources. At some point in time, a patient will lose connectivity while moving around the environment (e.g., leaving the nursing home or departing for a walk). Our system continuously monitors patients by adapting the ECG data rate to the expected movement and contact patterns. If the patient encounters a peer or smart device (e.g., PDAs, smartphones) that is a sink node with a direct connection to the internet, the system uses its peer as a mule to deliver its data to the data analysis center or hospital. During network partitions, the sample rate is reduced to maximize the amount of continuous data that can be stored.

Manuscript received on June 30, 2010; revised September 14, 2010; accepted October 6, 2010. Date of publication October 14, 2010; date of current version May 4, 2011. This work was supported in part by Microsoft Research, in part by the UCLA Wireless Health Institute, and in part by the Italian Ministry for Foreign Affairs under the program "Prosecuzione Laboratorio congiunto Italia-USA su reti wireless per cooperazione in mobilità".

T. Massey is with Johns Hopkins University Applied Physics Laboratory, Laurel, MD 20723 USA (e-mail: tammara.massey@jhuapl.edu).

G. Marfia is with the Dipartimento di Scienze dell'Informazione, Università di Bologna, Bologna 40126, Italy (e-mail: marfia@cs.unibo.it).

A. Stoelting is with FactSet Research Systems, Chicago, IL 60606 USA (e-mail: astoelting@factset.com).

R. Tomasi and M. A. Spirito are with the Pervasive Radio Technologies Laboratory, Istituto Superiore Mario Boella, Torino 10138, Italy (e-mail: tomasi@ismb.it; spirito@ismb.it).

M. Sarrafzadeh and G. Pau are with the Department of Computer Science, University of California, Los Angeles, CA 90095 USA (e-mail: majid@cs.ucla.edu; gpau@cs.ucla.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITB.2010.2087414

The proposed techniques specifically target an ECG sensor and the 2.4-Ghz bandwidth. However, these techniques can be applied to any high-data-rate embedded system, where quality data granularity is preferred. An outstanding challenge is how to balance the need for a high data rate to detect complex conditions, such as asymptomatic arrhythmia, and the need to have continuous data in order to diagnose infrequent conditions. The heart rate can diagnose some common conditions, such as tachycardia, bradycardia, and onset of change in patients. However, more complex conditions such as ventricular tachycardia or onset of change that focuses on a particular ventricle in the heart requires the entire ECG waveform. This research is applicable to mobile BSNs with embedded systems with limited memory capacity and high-data-rate sensors.

II. Related Work

Previous research explores techniques that opportunistic mobile systems use to deliver data to remote destinations. DataMule addresses the problem of sparse area networks by using mobile entities called mules to pick up data from static embedded systems and drop off the data at wired access points [1]. Message ferrying describes a proactive routing algorithm that allows embedded systems to deliver data on a specified route [2]. Another study, ZebraNet, uses a history-based protocol that intelligently selects nodes based on prior communication patterns and satisfies the tradeoffs between storage, bandwidth, and energy requirements [3]. Our social-network-analysis techniques manage the system's local resources so that the data granularity is modified to meet the storage restraints on the system.

Previous research in human mobility has used mathematical relationships to extract patterns in mobility. Previous work describes a statistical analysis of intercontact and intracontact periods in an opportunistic network [4]. Additionally, experimental studies have shown a correlation between line of sight and reception rate in mobile environments [5]. However, previous work has not leveraged these findings in the system design.

Research in online social networks has focused on how groups of users search, share, and organize content on online internet sites, such as Facebook and YouTube [6], [7]. In health care social networks, Merrill *et al.* analyzed patterns in public health networks to improve public health organizations and build collaborations [8]. This paper presents a *social system network* that focuses on communities or networks of individuals with lightweight medical systems that extracts patterns in connectivity, reception rate, location, and distance to improve system design. Specifically, an heuristic uses the estimated intracontact and intercontact time to provide continuous, high-quality data collection.

III. Overview

The physical size and cost-effectiveness of lightweight embedded systems result in limited processing power and communication bandwidth. Our goal is to create an infrastructure that manages the system's resources in order to lose as little data as possible. An overview of the techniques employed for mobile data collection of high-data-rate systems are shown in Fig. 1.

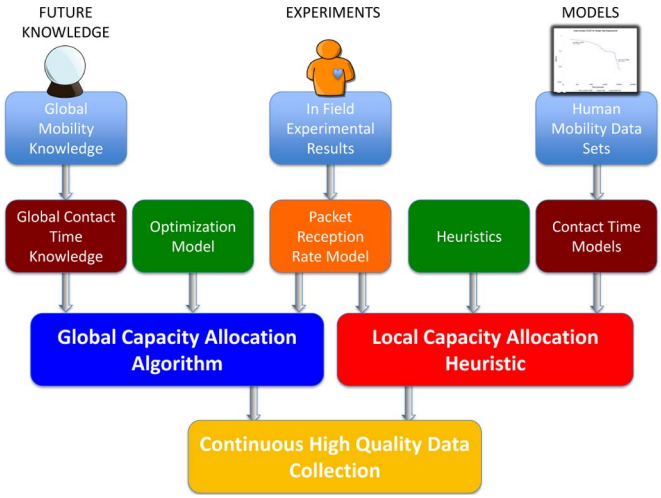

Fig. 1. Overview of methods employed for the continuous ECG data collection of mobile patients.

First, in field experiments analyzed the reception rate of embedded medical systems with various levels of visibility to a remote system. A statistical analysis of the observations is leveraged in a global algorithm and a local heuristic. A global capacity management algorithm was developed to determine the optimal granularity of data that should be collected on the system. A local heuristic exploited data on intercontact patterns and reception rates to efficiently manage the system's limited memory online. This paper compares the local heuristic and the optimal solution, and discusses the benefits of high-quality continuous data collection for social system networks.

IV. Social System Networks

Social system networks are a subset of social networks that specifically focus on techniques for mobile technologies. In social system networks, the interactions between systems in the network are an important channel for the transfer of data. We assume that the components of the social system network are interdependent and that patterns can be extracted from interactions. We also assume that the patterns observed in the networks are long lasting and that subgroups in these social system networks are likely to occur.

Although individuals are unique, human beings are creatures of habit whose regular behavioral patterns can be extracted and analyzed. All communities have common ties that draw individuals together into groups. Social system networks aim to efficiently and repeatedly extract the inherent periodicity in the networks of mobile systems. The interconnections, structures, dependencies, and patterns in the mobility of health systems are explored in social system networks.

The health care community has intrinsic interdependent relationships shaped by cyclic patterns in the service of care, routines in daily activities, and relationships between individuals. Social system networks of patients and health care providers (nurses, doctors, and emergency personnel) can opportunistically communicate with one another to upload the patient's data to a central server. These ubiquitous transport systems may

already be carried by health care providers in the form of cell phones or PDAs.

Similar to social networks, social system networks also leverage techniques in mathematics, such as statistics and graph theory. However, social systems networks can leverage the quantitative and ubiquitous nature of mobile systems to enable new finding in research areas that have only leveraged social networks. For example, researches in epidemics have observed how social networks aid in the spread of diseases. Social system network research can strengthen epidemic surveillance and preventive interventions by combining quantitative data collection of mobile systems with ubiquitous networks of information.

V. Experimentation

Over 24 h of experiments were performed to extract patterns in packet reception rate and signal strength from varying transmission ranges and node placements on the body. In these experiments, two people started on opposite ends of a 100-feet range. They walked toward each other, crossing at or near the midpoint of the range, to reach the other's starting point.

These experiments took place with two participants wearing embedded systems for communication nodes. Person A acted as the receiver and was fitted with six embedded systems. The six embedded systems were placed on person A's head, chest, arm, stomach, back, and leg. Person B acted as the transmitter and carried only a single node. However, the placement of the node on person B varied over the same six positions aforementioned. For each participant, six trials were performed along the 100-feet range with only the placement of B's transmitter on the body changing. Data on the reception rate and the distance between subjects was collected.

Results were measured by tracking the packet reception rate and received signal strength for each iteration of the experiment. The transmission rate was three packets per second with each packet carrying a sequence number resulted in approximately 10 800 samples being collected for each participant. Messaging was not reliable, therefore, the sequence number enabled a count of dropped packets to be taken.

Fig. 2 presents the results obtained from the experiments, in which the transmitter was mounted on the side, arm, or head. In these trials, signal propagation had a partial line of sight to the remote receiving nodes. As persons A and B move closer, the reception rate gradually increases. As the subjects move away from one another, the reception rate gradually decreases. The result is an approximately balanced reception rate curve peaking when the transmitters and receivers were in close proximity to one another.

Fig. 3 presents the results from the experiments, in which the transmitter was mounted on the front of the subject. These experiments allowed for a direct line of sight to the front mounted receiver nodes along the first half of the walking range. However, the line of sight was obstructed for the second half of the walking range. In these experiments, the reception rate is initially high, but drops drastically when the subjects pass one another and there is no longer a clear line of sight.

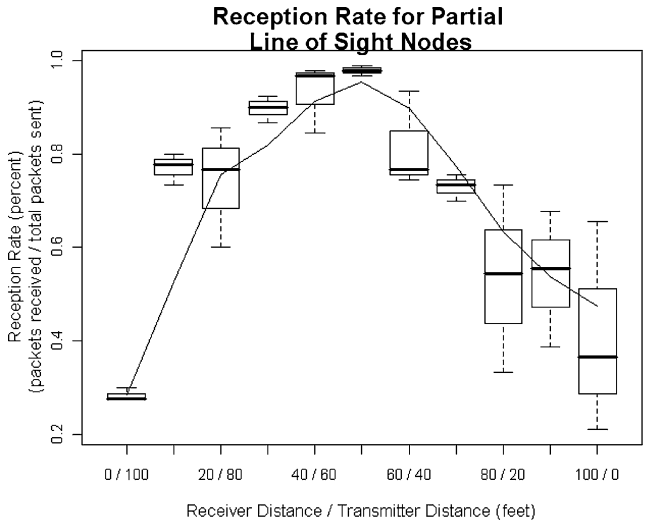

Fig. 2. Box plot of the reception rate as A and B move toward each other over a 100-feet range for nodes with partial line of sight placement (arm, ankle) with a trendline of a two-period moving average. The *x*-axis labeling shows A's and B's progress along the range, i.e., 20/80 shows that A and B have both moved 20 feet toward each other from their starting points.

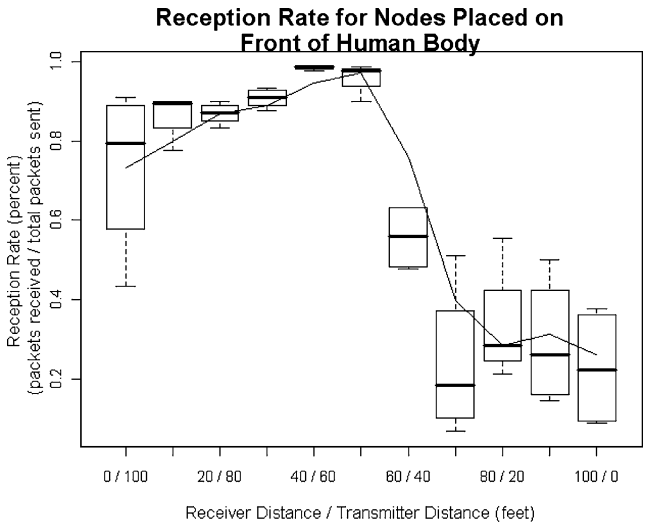

Fig. 3. Box plot of the reception rate as A and B move toward each other over a 100-feet range for nodes with direct/no direct line of sight place (chest, waist) with a trendline of the two-period moving average. The *x*-axis labeling shows A's and B's progress along the range, i.e., 20/80 shows that A and B have both moved 20 feet towards each other from their starting points.

In either scenario, the received signal strength is directly proportional to the proximity of the transmitter and receiver. The experiments: 1) determine how placement on the human body affects performance at various positions, and 2) identify patterns in reception rates, position, and distance.

VI. Health System Design

Architectural support in the hardware and software allow efficient data reconfiguration. The Continuous Health tELOS (Chelos) was built based on a two-led ECG embedded system developed by Fulford-Jones *et al.* [9]. The Chelos has a two-lead ECG, an MSP510 central processing unit, and a CC2420 radio

that communicates with 802.15.4. The system has a radio range of approximately 100 feet and consumes 23 mA at 9 V. The Chelos has 48 KB of ROM, 10 KB of RAM, and 1 MB of flash.

The circular first in first out (FIFO) buffer can store three levels of data granularity, 80, 40, and 1 Hz (heart rate). The circular buffer also has the functionality to store high priority data that will be accessed first. High priority data are transmitted with a flag to denote its importance. For high priority data, the pointer is moved backward and the data is stored before the normal data.

VII. Algorithms and Heuristics for Mobile Data Collection

In this mobile environment, patients walk around with embedded medical systems in an ad hoc network. When two nodes A and B meet, either A sends to B, B sends to A, or no transmission happens. In the case of no data transfer, the node's intercontact time is not stopped. The expected intercontact time is the expected delay between the two consecutive connections of a node with any sink node. The intracontact time is the length of the connection with the neighboring node.

The sink node is a node with connectivity to a storage server, typically the internet. If a system comes into contact with a sink node, it uploads its data. Often, the node does not have constant access to a sink node. In this case, the embedded system must manage its resources to prevent data loss and pass data through its neighbors in a peer-to-peer fashion whenever it encounters another node.

A. Global Algorithm for Capacity Allocation

The capacity allocation problem can be modeled as a budgeting problem. Let $V = v_1, ..., v_N$ be the nodes in the network. Communication links in the network are organized as a graph $G(V, E)$, where V is the set of nodes and E is a set of edges. An edge (v_i, v_j) indicates the existence of a communication link from v_i to v_j.

Definition: A valid path P is a sequence of nodes $P = < v_1, ..., v_S >$ such that there is a link from v_i to v_j, where $1 < i, j < S$. The start node is the designated source and the end node is the designated destination, the sink S. Data traveling to the sink node flows through upstream nodes, u_i.

A link in this network is denoted $L(v_i, v_j)$, and it has an associated link capacity of how much data can be sent across link v_i to v_j. This link capacity is referred to as C.

The buffer of m_i, is the amount of memory at node$_i$, where i ranges from 1 to N. N is the number of nodes in the network. Each buffer can sample data at several different granularities g_i, where i ranges from 1 to M.

Problem Statement: Given a network, assign data granularity so that each node maximizes the amount of data received at the sink d_{iSk} and the data granularity g. The objective function also minimizes the amount of data lost l, as shown in the following:

Problem formulation:

$$\text{Maximize:} \sum_{i=1}^{N} \sum_{k=1}^{T} \left(d_{iSk} + \left(\sum_{w=1}^{W} g_{ikw} \right) - \sum_{i=1}^{N} l_{ik} \right). \quad (1)$$

Subject to the following constraints:

$$m_{ik} \leq C \qquad \forall i \ \forall k \quad (2)$$

$$m_{ik} \geq 0 \qquad \forall i \ \forall k \quad (3)$$

$$m_{sk} < \infty \qquad \forall k \quad (4)$$

$$\sum_{w=1}^{W} g_{ikw} = 1 \qquad \forall i \ \forall k \quad (5)$$

$$G_{ikw} = S_w(E_{ik} + A_{ik}) \qquad \forall i \ \forall k \ \forall w \quad (6)$$

$$m_{i(k+1)} = m_{ik} + \sum_{w=1}^{W} (G_{ikw} g_{ikw}) + \sum_{i=1} u_{ijk} - d_{iSk} - l_{ik} \qquad \forall i \ \forall k \quad (7)$$

$$d_{ijk} <= R_{ijk} A_{ijk} \qquad \forall i \ \forall j \ \forall k. \quad (8)$$

Constraints (2) and (3) guarantees that the memory or buffer on each embedded system m does not exceed its capacity C. Constraint (4) satisfies the fact that the capacity of the buffer at the sink node m_{sk} is less than infinity. g is the percent of the contact period that the node is at a particular data granularity. In our deployment, the sample rates were 80, 40, and 1 Hz (heart rate). Constraint (5) ensures that these granularity ratios g for the various sample rates equals 1. Equation (6) calculates the newly generated data, G. G is equal to the data size S times the inter-contact time E and the intracontact time A. Constraint (7) formulates that the buffer at $k + 1$ is equal to summation of the node's current buffer (m_{ik}), newly generated data G for each sample granularity ratio g, and the data from upstream nodes u. Also, the data sent d and data lost l reduces the buffer size. Constraint (8) defines that the data transmitted d is less than or equal to the reception rate R times the intracontact time A. The reception rate R is a metric that used the probability of packets received and packets sent to estimate the link quality. Linear programming was used to solve the constraints using Cplex, an optimization software package.

The application of nonparametric statistical techniques to represent the stochastic nature of the packets was used in the problem formulation. MLE was used to determine the distribution of reception rates gathered from the measurement studies. MLE is a statistical method that fits data to a statistical model. In this case, we fit the data to the model of the distribution of reception rate in the experimental data (see Fig. 4). Systems placed on the side with a partial line of sight have a normal distribution with the peak reception rate of systems occurring when the transmitter and receiver are closest to each other. On the other hand, systems placed on the front have a skewed distribution with the highest reception rates occurring in the first half of the experiment range. The skewed distribution stems from a direct line of sight to the remote system at the beginning of the experiment and no line of sight at the end of the experiment. The MLE represents the stochastic nature of the data by modeling the reception rates in the same distribution as the data.

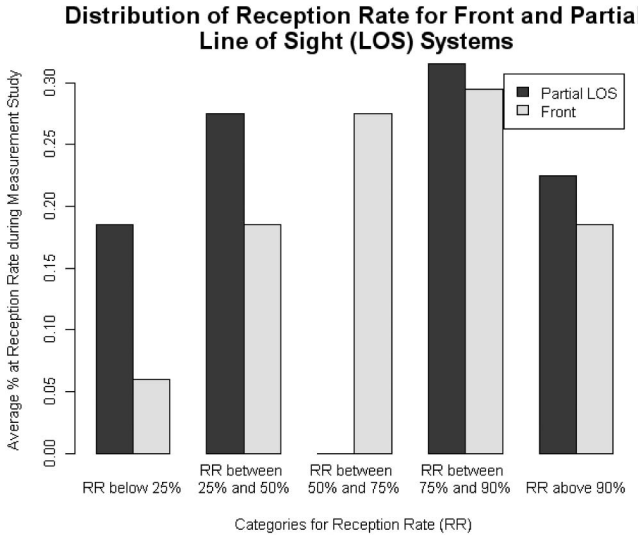

Fig. 4. Distribution of reception rate (RR) based on node placement on the body. RR is the number of packets received divided by the total packets sent.

B. Local Heuristic for Capacity Allocation

A local heuristic was used on the embedded system. Three factors were estimated on the embedded system: the intercontact time, the intracontact time, and the location of system on the human body. The probability of experiencing any given intercontact time was derived from a power law distribution. The complementary cumulative distribution function (CCDF) of the intercontact times is used to calculate the power law distribution. Previous research demonstrates that most mobility models follow a power law distribution with an exponential tail for the intercontact times between mobile devices [10]. The estimated intercontact and intracontact times is used to determine the estimated gaps between the communication and the length of connections. The location of the system aids in estimating the reception rate to the remote system.

In the following is the equation for the power law distribution. $P(t)$ is the expected probability of an interconnect period being less than or equal to time t. x is the current intercontact delay. y is the scaling variable that preserves the proportionality of the scale change and the shape of the function. y should always be less than 1. Previous work established that there are power law mobility patterns in humans, where y changes depending on the type of environment (conference, laboratory, etc), wireless communication (Bluetooth, WIFI, etc), and the number of participants [4]. For the default scaling factor, we used the scaling factor that was similar to our deployment in terms of the number of participants and environment [11]–[14], 0.59

$$P(t) = x \times t^{-y}. \tag{9}$$

In (9), the current time t is the time elapsed since the last contact. $P(t)$ gives the probability of a contact happening within the current time range t. Also, the probability indicates the percent of intercontact times that are longer than the current intercontact period experienced thus far. Relatively short intercontact times are most likely to occur over the course of the deployment. However, significantly long intercontact times will also occur.

The dataset has a power law distribution with an exponential tail for the intercontact time. However, the exponential tail begins after the sample rate would drop to 1. Therefore, it is unnecessary to add the exponential decay to the heuristic. A best fit linear regression line for the power law equation for the dataset, where x is 15.9 and y is 0.59, is 0.99.

The next intercontact time as a percentage of the remaining probability range was run on a deployment dataset. For the Chelos embedded medical system, between 0.3% and $5.2\%$of the overall data was dropped with approximation factors ranging between 0.1 and 0.9. An approximation factor of 0.5 was chosen because it resulted in a loss of 1% of the total data collected.

In the case of intracontact time, the exponential decay is important. When the time is less than 3000 s, x is 33.5 and y is 0.7 with a best fit linear regression line of 0.97. When the time is greater than 3000 s, x is 31145 and y is 1.56 with a best fit linear regression line of 0.97. Additionally, the placement of the embedded system affects the reception rate and the probability that the intracontact time will last long enough to send all of the accumulated data. Patterns in the reception rate can be used to draw conclusions about a system's position on the body in the local heuristic. Throughout the duration of a contact period, the sequence numbers of incoming packets are monitored. Two counts are kept, an expected packet count and an actual packet count. Whenever a sequence number is skipped in the incoming packet stream, the expected packet count reflects the lost packets. Comparing the loss rates on either side of the signal strength maximum yields clues to the line of sight experienced between transmitter and receiver.

The packet reception rate over the length of the connection (intracontact period) before and after the peak in signal strength was analyzed to determine system placement on the body. The reception rates were at its peak when the systems were in close proximity with a clear line of sight to the other system in the experimental measurements. After the connection has been lost, the reception rate statistics for both halves can be calculated and compared. This rate is found by simply dividing the number of packets received by the number of total packets expected for each half. If the reception rate before the peak was significantly higher than that after the peak, the system usually experienced a clear line of sight to the transmitter before the peak and no line of sight after the peak. A significant difference in reception rates was when the absolute difference was greater than 15%. Likewise, for a lower reception rate followed by a higher reception rate, the system was usually initially obstructed followed by a line of sight to the embedded system. When the reception rates were relatively balanced, the system had a consistent level of interference the entire time, i.e., the system continuously had partial line of sight to the remote system during the connection.

VIII. Results and Deployment

The local heuristic presented previously only has the knowledge of the present, the past, its own system resources, and the resources of recent neighbors. The local heuristic was compared

with the global algorithm that has knowledge of its expected intercontact delay to encounter its next neighbor, available buffer, and its expected delay to encounter a sink node. The expected intercontact delay in particular determines the granularity of sampling that the embedded system will use. In contrast to the global algorithm, the local algorithm estimates the intercontact time through history-based techniques. The global algorithm determines the optimal sampling rate for each node in order to get the maximum data to the sink.

The global algorithm and local heuristic is applied to each snapshot of the corresponding dataset. The sample rate for each node is calculated with the local heuristic that leverages the expected intracontact time, the expected intercontact time, and the system position metric presented earlier in the paper. We assume that the position of the node in reference to its receiving node does not change more than once every 10 min for deployment 1 at Intel and deployment 2 at Cambridge. In deployment 3 at UCLA, the system's position was calculated at the end of every intracontact period.

A three-day deployment was conducted to analyze the accuracy of the online local heuristic. Six participants each wore a transmitter programmed with our algorithms described previously, and a single base station was used. All six participants belonged to the same laboratory, where the one base station was located. However, only three of the six participants spent much of their day there. The other three would pass through sporadically throughout the duration of the experiment.

Upon contact with the sink node, the node would send the estimated system position, sequence number, and data collected since the last contact with the sink. The mobile position heuristic accurately determined online the position of the embedded system 93% of the time. The accuracy of determining the online position was calculated by comparing the actual positions of the systems on the body as reported by the participants with the results of the positioning algorithm that were sent with each packet. The error in the positioning algorithm is most likely attributed to long seated positions in proximity to the base station, where the overall reception rate was approximately even during the entire transmission. As noted previously, proximity and obstruction are the dominant factors in signal strength. Any prolonged exposure leading to an incorrect positioning estimation is offset by the fact that a connection to the sink is open and data are being delivered.

From the time stamps gathered on the incoming packets, we can determine the length of the intercontact and intracontact periods. From these periods, we can generate a histogram to observe the individual frequencies. Duration categories for this experiment are divided into 30 s intervals. After accounting for each intercontact and intracontact periods in the correct category, a CCDF can be generated. Figs. 5 and 6 show the CCDFs for the intercontact and intracontact periods, and the corresponding best fit lines. The exponent in the power law equation is slightly lower than previous deployments with similar characteristics.

Experiments were run on the deployment dataset of human mobility to explore how the global and local heuristic compared. Additionally, our online heuristic and global algorithm

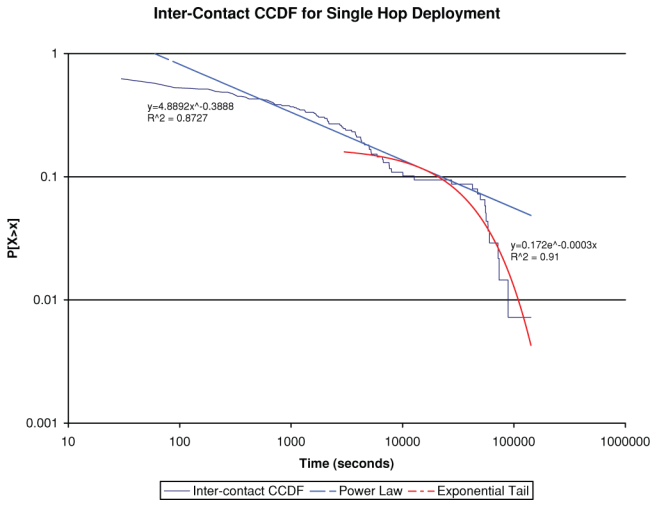

Fig. 5. Intercontact CCDF of deployment.

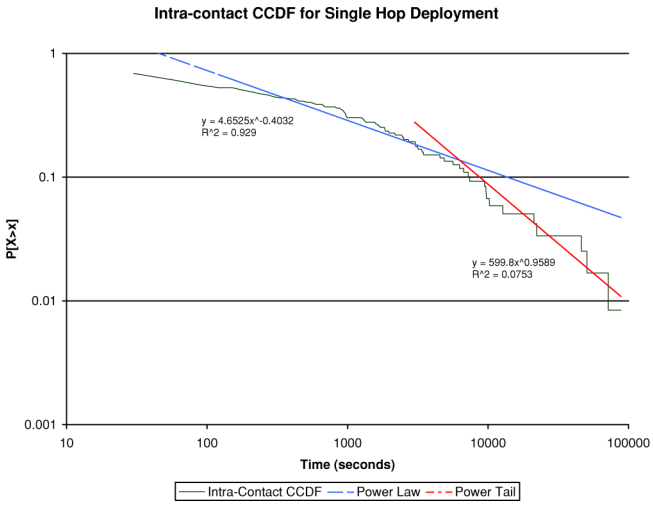

Fig. 6. Intracontact CCDF of deployment.

was run on several mobility datasets [11], [15]. While the mobility datasets for deployments 1 and 2 gave accurate information on human mobility, they lacked two main properties needed for our study. First, the datasets lacked any reference to a buffer limitation and its implications on the embedded system. Second, the dataset did not record the link quality of the links that it observed. The datasets only observed intercontact and intracontact times. For the first parameter, we used the hardware limitations of the Chelos device, 48 Kb ROM, 10 Kb RAM, and 1 MB of flash memory. For the second parameter, we used the measurement study done on the 802.15.4 radio. The data granularity of the dataset was every 10 min. An assumption that was made was that those links were stable during those 10 min and that no other nodes were in the vicinity between measurements.

The results of the deployment are shown below in Fig. 7, where the global algorithm is compared to a local solution. Our results show that our heuristic is close to the global solution.

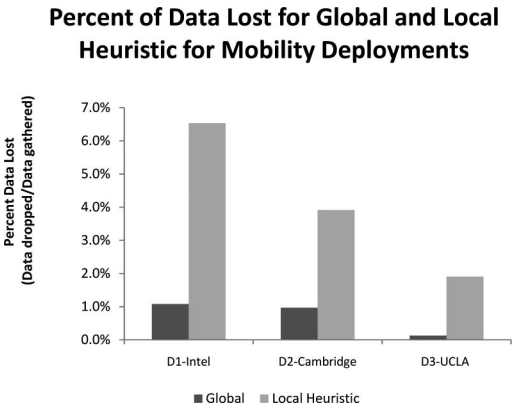

Fig. 7. Data lost of the naive, heuristic, and global solution on different datasets.

Unfortunately, not even the global solution can prevent all data loss. Notice that in some cases even at the lowest sampling rate, data would be lost due to the period of the network partition. We also compare the heuristic to several other datasets in Fig. 7 [11], [15]. Our online heuristic performs well on other mobility datasets and the amount of data loss compared with data collected does not rise over 7%.

IX. Discussion

Our research is the first step in providing a reliable solution to continuous data collection of high-data-rate systems. We specifically analyzed how the reception rate varies in a mobile environment and addressed this factor in our research. However, in order to build a comprehensive solution, all factors that affect the system should be addressed. Open challenges and assumptions in the system, deployment, and results must be addressed for a functioning BSN that operates as expected in a dynamic environment.

In terms of the system, we did not address security of the data on the embedded system. The security requirements for the transmission of data among mobile health systems are a topic that should be addressed in the future. Health systems pose stringent requirements in terms of data confidentiality. Researchers have presented security solutions for peer-to-peer and ad hoc networks. However, experimental deployments similar to our study have rarely explored with mobile security solutions. The implications of additional security on other aspects of the system outside of simulation should be explored. For example, an additional delay due to additional encryption processing will affect how much data can be sent during a time period.

Also, our research also made several assumptions in regards to the deployment. Our research assumed that the patterns observed in the social system network are long lasting. However, patterns in social networks often vary and drift over time. Even in our analysis, the scaling factor varied depending on the type of deployment (environment, types of participants, etc.). It would be beneficial to compute the scaling factor online. Online computation of the scaling factor and other history-based variables are a challenge due to the limited storage capacity on embedded systems. Techniques to compute the scaling factor online without storing large amount of data should be explored.

Additionally, we simplified the problem by assuming single-hop networking, a small network of peers, and a three-day deployment. Our current global algorithm provides the capability for multi-hop routing. In the future, we would like to analyze the performance of the heuristic and global solution using multi-hop routing algorithms. Moreover, a small network of six participants carried our systems for three days. With improved charging capabilities and more resources, we would like to run experiments for weeks and months in order to get more accurate data on how well the heuristic performs over a longer period of time.

In order for ubiquitous data collection to occur, these health systems must be part of a system that the user has a strong incentive to charge and continuously carry with them. The design of multi-platform systems, such as a medical cell phone that address multiple needs of the users should be explored. Multisensor platforms that have resources for other purposes can possibly improve the algorithm. For example, locality information from GPS or cellular towers, may be leveraged to estimate the intercontact and intracontact times.

Finally, even though the analysis of the data gathered at the data analysis center is out of scope of this paper, it is an important factor that should also be addressed. The development of efficient algorithms at data centers to make sense of the multitude of data is essential in building systems that accurately notify users of abnormalities.

X. Conclusion

The efficient use of limited resources can be obtained through techniques that collect continuous data from a patient despite long intercontact periods. In this paper, our social system network analysis techniques allow the medical system to mitigate data loss by modifying the sampling rate and leveraging patterns in human mobility and reception rate in a BSN. We incorporated the stochastic nature of the data from experimental studies on human mobility in a global algorithm to determine the optimal data granularity. Additionally, we presented an online heuristic for lightweight health systems that mitigates partitions in communication by leveraging social patterns in reception rate, intercontact time, and intracontact time.

Acknowledgment

The authors would like to thank Prof. M. Orefice and Prof. G. Borio of the Laboratory of Antennas and EMC (LACE), Politecnico di Torino. Experimental measurements were performed in Torino and Vercelli (Italy).

References

[1] R. Shah, S. Roy, S. Jain, and W. Brunette, "Data mules: Modeling a three-tier architecture for sparse sensor networks," in *Proc. 1st IEEE Int. Workshop Sens. Netw. Protocols Appl. (SNPA)*, 2003, pp. 30–41.

[2] W. Zhao and M. H. Ammar, "Message ferrying: Proactive routing in highly-partitioned wireless ad hoc networks," in *Proc. 9th IEEE Workshop Future Trends Distrib. Comput. Syst. (FTDCS)*, May 2003, pp. 308–314.

[3] P. Juang, H. Oki, Y. Wang, M. Martonosi, L. S. Peh, and D. Rubenstein, "Energy-efficient computing for wildlife tracking: Design tradeoffs and early experiences with zebranet," *SIGARCH Comput. Archit. News*, vol. 30, no. 5, pp. 96–107, 2002.
[4] A. Chaintreau, P. Hui, C. Diot, R. Gass, and J. Scott, "Impact of human mobility on opportunistic forwarding algorithms," *IEEE Trans. Mobile Comput.*, vol. 6, no. 6, pp. 606–620, Jun. 2007.
[5] T. Massey, G. Marfia, M. Potkonjak, and M. Sarrafzadeh, "Experimental analysis of a mobile health system for mood disorders," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 241–247, Mar. 2010.
[6] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in *Proc. 7th ACM SIGCOMM Conf. Internet Meas.*, 2007, pp. 29–42.
[7] N. Ellison, C. Steinfield, and C. Lampe, "The benefits of facebook 'friends': Social capital and college students' use of online social network sites," *J. Comput.-Mediated Commun.*, vol. 12, no. 4, pp. 1143–1168, 2007.
[8] J. Merrill, M. Rockoff, S. Bakken, and K. Carley, "Organizational network analysis: A method to model information in public health work," in *Proc. Amer. Med. Informatics Assoc.*, 2006, p. 1030.
[9] T. Fulford-Jones, G.-Y. Wei, and M. Welsh, "A portable, low-power, wireless two-lead EKG system," in *Proc. IEEE Eng. Med. Biol. Soc.*, 2004, pp. 2141–2144.
[10] T. Karagiannis, J.-Y. Le Boudec, and M. Vojnovi, "Power law and exponential decay of inter contact times between mobile devices," in *Proc. 13th Annu. ACM Int. Conf. Mobile Comput. Netw. (MobiCom)*, New York: ACM, 2007, pp. 183–194.
[11] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot, "Pocket switched networks and human mobility in conference environments," in *Proc. ACM SIGCOMM Workshop Delay-Tolerant Netw. (WDTN)*, New York: ACM, 2005, pp. 244–251.
[12] M. McNett and G. M. Voelker, "Access and mobility of wireless PDA users," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 9, no. 2, pp. 40–55, 2005.
[13] T. Henderson, D. Kotz, and I. Abyzov, "The changing usage of a mature campus-wide wireless network," in *Proc. 10th Annu. Int. Conf. Mobile Comput. Netw (MobiCom)*, New York: ACM, 2004, pp. 187–201.
[14] N. Eagle and A. S. Pentland, "Reality mining: Sensing complex social systems," *Pers. Ubiquitous Comput.*, vol. 10, no. 4, pp. 255–268, 2006.
[15] J. Scott, R. Gass, J. Crowcroft, P. Hui, C. Diot, and A. Chaintreau. (2006, Jan.). *CRAWDAD data set cambridge/haggle (v. 2006-01-31)* Downloaded from [Online]. Available: http://crawdad.cs.dartmouth.edu/cambridge/haggle.

Tammara Massey (M'04) received the M.S. degree in computer science from the Georgia Institute of Technology, Atlanta, in 2004, and the Ph.D. degree in computer science from the University of California, Los Angeles, in 2009.

In 2009, she joined the Johns Hopkins University Applied Physics Laboratory, Laurel, MD, as a Systems Engineer and Researcher. Her research interests include mobile health systems, medical informatics, biometrics, sensor-enabled embedded systems, social system networks, and preventive interventions.

Gustavo Marfia received the Laurea degree in telecommunications engineering from the Università di Pisa, Italy, in 2003, and the Ph.D. degree in computer science from the University of California, Los Angeles, in 2009.

He is currently a Postdoctoral Research Fellow at the Dipartimento di Scienze dell'Informazione, University of Bologna, Bologna, Italy. His research interests include ad hoc, vehicular, and biomedical networks.

Adam Stoelting received the M.S. degree in computer science from the University of California, Los Angeles, in 2009.

He is currently a Software Engineer at FactSet Research Systems. His research interest includes distributed systems development and scheduling.

Riccardo Tomasi received the double M.S. degree in telecommunications engineering from the Politecnico di Torino,Torino, Italy, and the Institut National Polytechnique de Grenoble, Grenoble, France.

He is currently a Researcher at the Pervasive Radio Technologies Laboratory, Istituto Superiore Mario Boella, Turin, Italy. His research interests include wireless sensor networks, with a focus on protocols and software tools to support development and integration of such technologies within the internet of things.

Maurizio A. Spirito (S'98–M'01) received the M.S. degree in electronics engineering and the Ph.D. degree in electronics and telecommunications engineering from the Politecnico di Torino, Torino, Italy, in 1996 and 1999, respectively.

From 1998 to 2003, he was with the Nokia Research Center, Helsinki, Finland, as a Senior Research Engineer and Project Manager. In 2003, he joined Istituto Superiore Mario Boella, Torino, Italy, to head the Pervasive Radio Technologies Laboratory. His current research interests include short range wireless communications technologies and applications for the internet of things.

Majid Sarrafzadeh (M'87–SM'92–F'96) received the B.S., M.S., and Ph.D. degrees in electrical and computer engineering from the University of Illinois at Urbana-Champaign, IL, in 1982, 1984, and 1987, respectively .

In 1987, he joined Northwestern University as an Assistant Professor. Since 2000, he has been with the Department of Computer Science, University of California, Los Angeles. His research interests include embedded and reconfigurable computing, very large scale integration CAD, and the design and analysis of algorithms.

Giovanni Pau received the Laurea Doctorate in computer science and the Ph.D. degree in computer engineering from the University of Bologna, Bologna, Italy.

He is currently at the Department of Computer Science, University of California, Los Angeles, as a Research Scientist. His research interests include wireless ad hoc networks, urban sensing, and distributed computing. He is the architect of the UCLA Campus Vehicular Testbed, an experimental facility to support urban sensing for pollution monitoring, and intelligent transportation systems.