Opportunistic Energy-Efficient Contact Probing in Delay-Tolerant Applications

Wei Wang, *Member, IEEE*, Mehul Motani, *Member, IEEE*, and Vikram Srinivasan, *Member, IEEE*

***Abstract*—In many delay-tolerant applications, information is opportunistically exchanged between mobile devices that encounter each other. In order to affect such information exchange, mobile devices must have knowledge of other devices in their vicinity. We consider scenarios in which there is no infrastructure and devices must probe their environment to discover other devices. This can be an extremely energy-consuming process and highlights the need for energy-conscious contact-probing mechanisms. If devices probe very infrequently, they might miss many of their contacts. On the other hand, frequent contact probing might be energy inefficient. In this paper, we investigate the tradeoff between the probability of missing a contact and the contact-probing frequency. First, via theoretical analysis, we characterize the tradeoff between the probability of a missed contact and the contact-probing interval for stationary processes. Next, for time-varying contact arrival rates, we provide an optimization framework to compute the optimal contact-probing interval as a function of the arrival rate. We characterize real-world contact patterns via Bluetooth phone contact-logging experiments and show that the contact arrival process is self-similar. We design STAR, a contact-probing algorithm that adapts to the contact arrival process. Instead of using constant probing intervals, STAR dynamically chooses the probing interval using both the short-term contact history and the long-term history based on time of day information. Via trace-driven simulations on our experimental data, we demonstrate that STAR requires three to five times less energy for device discovery than a constant contact-probing interval scheme.**

***Index Terms*—Bluetooth, delay-tolerant networking (DTN), energy efficiency.**

I. Introduction

SINCE its inception, the goal of networking research has been to provide instant, anytime, anywhere access to information. However, in recent times, research interest has been piqued by a new class of applications that are tolerant to delay. In several of these applications, information is exchanged opportunistically between devices when they are within communication range of each other. In other words, information transport is governed by the mobility of information carriers and their underlying contact patterns.

The notion of delay tolerance is useful in a variety of scenarios. One compelling application is providing connectivity and network services during disasters and in rural environments, where network infrastructure is minimal or nonexistent. Another example comes from software developers who are developing dating applications for mobile phones. The profile of an ideal partner is entered into a Bluetooth-based mobile phone, which alerts the user whenever a matching profile is detected in the vicinity (e.g., www.bedd.com).

Current research and development efforts for delay-tolerant applications fall broadly into two categories: delay-tolerant networking (DTN) [2] and delay-tolerant databases (DTD). For DTN applications, the goal is to enable communication between specific source–destination pairs in the network. Research in this area has involved studying algorithmic issues such as routing in networks [3], fundamental issues such as scaling laws [4], and performance bounds of routing algorithms [5] based on real-world contact patterns.

DTD applications have been driven by the observation that mobile devices are becoming increasingly powerful in terms of computation and storage and have multiple radio interfaces such as Bluetooth, 3G, WiFi, etc. [6]. An effort is also being made by phone manufacturers to embed sensors in these phones to acquire and store personal information (health-related) and for environmental monitoring [7]. As a consequence, these devices store large volumes of digital information such as songs, photographs, and sensory data and constitute a distributed geographic database. The dating application stated earlier is an example of a DTD application. The research community has investigated opportunistic query propagation and data aggregation algorithms, based on device proximity, in [6] and [8]–[10].

For both DTN and DTD applications, the common fundamental primitive is the opportunistic exchange of information between mobile devices when they are in communication range of each other. In order to enable such exchanges, devices will have to constantly probe the environment to discover other devices in the vicinity. Not surprisingly, device discovery[1] is an extremely energy-consuming process. To understand this better, we made measurements on a Nokia 6600 mobile phone. The current drawn was: 1) 38.61 mA for Bluetooth device discovery; 2) 9.33 mA for enabling the device to be discoverable; 3) 51.47 mA for Bluetooth data transfer; and 4) 38.68 mA for making a phone call. In other words, the device discovery process is as energy-intensive as making a phone call.

Our measurements clearly motivate the need for energy-conscious device-discovery algorithms. Although the measurements in this paper are based on Bluetooth devices, our

Manuscript received November 25, 2007; revised August 12, 2008; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor B. Levine. First published June 30, 2009; current version published October 14, 2009. Part of this work was presented at ACM MobiCom 2007, Montreal, QC, Canada.

W. Wang was with the National University of Singapore, Singapore 119620, Singapore. He is now with Microsoft Research Asia, Beijing 100190, China (e-mail: wei.wang@microsoft.com; wangwei.ww@gmail.com).

M. Motani is with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore 119620, Singapore (e-mail: motani@nus.edu.sg).

V. Srinivasan is with Bell Labs Research, Bangalore 560095, India (e-mail: vikramsr@alcatel-lucent.com).

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/TNET.2008.2008990

[1]We use device discovery and contact probing interchangeably in this paper.

device-discovery model can be used in other protocols, such as 802.11 or UWB. As there are no centralized coordinators in DTNs, it is hard to synchronize devices in the network, and devices normally do not have information on which channel should be used in device discovery. In device-discovery protocols used in DTN, a mobile may need to send multiple packets over a number of channels in order to scan for devices around it. Therefore, the high probing energy consumption could be a common problem in delay-tolerant networks.

One strategy to conserve energy is to increase the time between subsequent device discoveries. The consequence of this is that devices in the vicinity may go undiscovered, and opportunities to exchange data are lost. This points to a tradeoff between energy and missed opportunities. For strategies that use a constant device-discovery interval, the larger the probing interval, the larger the missed opportunities and vice-versa. However, in stochastic environments, device discovery should be done adaptively by choosing the probing interval based on the state of the environment. For example, late at night at home, device discovery can occur at large intervals without missing many contacts, while on the subway to work, device discovery should be done more frequently to catch the myriad of new contacts.

In this paper, we investigate the design of energy-conscious, adaptive contact-probing algorithms that trade off energy consumption and the probability of missing a contact. Specifically, we make the following contributions.

A. Theoretical Framework

We first develop a theoretical framework to characterize the tradeoff between the average contact-probing interval T and the contact missing probability P_{miss}. We show that if the contact duration distribution is independent and identically distributed (*i.i.d.*) and the contact arrival rate is constant, then for a given missing probability constraint, the optimal contact-probing interval is 1) constant and 2) depends only on contact duration distribution, independent of the contact arrival distribution. When the contact arrival rate is time varying, the optimal contact-probing interval is a function of the contact arrival rate. We provide an optimization framework to compute the optimal contact-probing interval as a function of the contact arrival rate. This theoretical base provides us with bounds on performance and also aids in the investigation and design of adaptive contact-probing algorithms. See Sections II and III.

B. Real-World Contact Pattern Experiments and Analysis

The theoretical development above depends on the contact pattern statistics. To understand real-world contact patterns, we conducted a large-scale data logging experiment [11]. Nine volunteers were given Bluetooth devices equipped with a software program that probed for contacts every 30 s and logged information about other Bluetooth devices that came within range. Our database contains the largest number of unique devices discovered, compared to existing work [12]–[14]. We conducted rigorous analysis on the data. We confirmed that the contact duration is Pareto-distributed. Moreover, our data analysis indicates weak correlations in contact patterns at 24-h time lags. Finally, we show that the contact process is self-similar with a Hurst parameter of 0.76. See Section IV.

C. Algorithm Design and Validation

Finally, using insights gleaned from the theory and the data analysis, we investigate adaptive contact-probing algorithms. We design STAR, an algorithm that strives to estimate the arrival rate and adapt the contact-probing interval based on this estimate. We consider using several estimation methods in conjunction with STAR and compare them using trace-driven simulations. We show that a simple weighted average of previous arrival rate estimates (STAR-PTS) works well and has 1/3 the energy consumption of the constant nonadaptive contact-probing algorithm for a missing probability of 0.2. The minimum mean squared error estimator approach (STAR-MMSE) is even better and has 1/5 the energy consumption of the nonadaptive scheme. See Sections V and VI.

II. Modeling the Contact Process

A. System Model and Assumptions

Assume that every device probes the environment governed by some contact-probing algorithm. When a device *A* probes its environment, all devices that hear the probe respond to device *A* with some information (e.g., identity, services available, etc.). Based on this information, *A* may choose to exchange data with some of its neighbors. We define two device,s *A* and *B*, to be in *contact* if they are within communication range of each other[2]. The duration over which devices *A* and *B* are continuously in contact is called the *contact duration* for that contact. If neither device probes its environment during the contact duration, then we have a *missed contact*. We further assume, for the sake of analysis, that each probe is an impulse and does not consume any time.

We assume that each probe consumes the same amount of energy of e_p. When the average contact-probing interval is T, the energy consumption rate of the device will be e_p/T. With the same number of missed contacts, the algorithm that uses fewer probes (longer average probing interval) will be more energy efficient.

For a given device, we assume that the contact durations $\mathbf{t}_D(i)$ are *i.i.d.* random variables with cumulative distribution function (cdf) of $F_D(x)$ and mean $\mathbb{E}\{\mathbf{t}_D\} = 1/\mu$.

We assume that the contact arrival process is stationary. The contact arrival process is characterized by the *intercontact time* $\mathbf{t}_C(i)$, defined as the time between the ith and $(i+1)$-th contacts. The stationary assumption is that the sequence of *intercontact times* is a wide-sense stationary process; i.e., the $\mathbf{t}_C(i)$ have a constant mean of $\mathbb{E}\{\mathbf{t}_C\} = 1/\lambda$. We relax this assumption and deal with nonstationary contact arrival processes in the next section.

See Fig. 1 for an illustration of contact duration $\mathbf{t}_D(i)$ and intercontact time $\mathbf{t}_C(i)$.

B. Missing Probability

The missing probability P_{miss} is the probability a contact that occurs is not detected. For the following analysis, we assume

[2]We do not assume a perfect communication region here. The imperfectness in communication is embedded in our contact data gathering process that uses real devices.

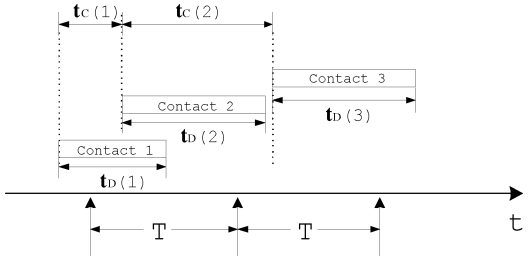


Fig. 1. Illustrating the contacts for a specific user probing at a constant interval T.

that for a device *A*, a contact *B* is missed if *B* is not discovered by *A*'s probes. We will relax this later to compute the missing probability when neither *A* nor *B*'s probes discover each other. Let us first consider the simplest possible contact strategy, where each device probes for contacts at constant intervals of T (see Fig. 1).

We first consider the case when contact probes are reliable; i.e., every probe can successfully detect all contacts around the device. If $\mathbf{t}_D > T$, the contact will never be missed. Consider a contact i that is initiated at time $[nT - x, nT - x + dx]$, where dx is an arbitrarily small value so that the time interval can be treated as a time point. This contact will not be detected by the mobile at time nT if $\mathbf{t}_D(i) < x$, which happens with probability of $F_D(x)$. By Blackwell's Theorem in renewal theory [15], when the contact process is a general renewal process with a nonlattice distribution function, we can calculate the long-term average missing probability given the probing interval of T as

$$\lim_{n\to\infty} P_{\text{miss}}(T) = \frac{1}{T}\int_0^T F_D(x)dx. \tag{1}$$

Note that we do not need to restrict the intercontact time distribution to be memoryless to perform the averaging procedure in (1).

Consider the mean value of $\mathbf{t}_D$ given that $\mathbf{t}_D < T$, which is $\mathbb{E}\{\mathbf{t}_D|\mathbf{t}_D < T\} = \int_0^T [1 - F_D(x)/F_D(T)]dx$. Then, $P_{\text{miss}}(T)$ can be expressed as

$$P_{\text{miss}}(T) = \frac{T - \mathbb{E}\{\mathbf{t}_D|\mathbf{t}_D < T\}}{T} \times F_D(T). \tag{2}$$

For a constant contact-probing interval of T, any contact with duration larger than T will always be detected. Equation (2) shows that for a contact with duration shorter than T, the expected probability that the contact will be missed is $\frac{T-\mathbb{E}\{\mathbf{t}_D|\mathbf{t}_D<T\}}{T}$. Therefore, only contacts with duration smaller than T will be missed. If all the contacts with $\mathbf{t}_D < T$ have zero contact duration, then $P_{\text{miss}}(T)$ will be exactly $F_D(T)$.

The key observation here is that for a constant probing scheme, if the contact durations are *i.i.d.* and the intercontact times are stationary, then the missing probability depends only on the distribution of the contact duration and the contact-probing interval T. It is independent of the intercontact time distribution.

In real networks, contact probes are unreliable due to the nondeterministic nature of wireless channel. Consider the case where a probe can miss a present contact with probability of κ. The overall missing probability P_{miss} can be derived in a similar way (see the Appendix). Similar to the reliable probes case, the missing probability for unreliable probes is also independent of the intercontact time distribution. In the rest of this paper, we only consider cases where all contact probes are reliable.

C. Optimal Contact-Probing Scheme

We will now prove the following theorem about the constant contact-probing scheme described earlier.

Theorem 1: Consider a contact process for which the contact durations are *i.i.d* and the distribution of intercontact times is stationary, with an expected intercontact time of $\frac{1}{\lambda}$. Consider the class of contact-probing strategies that do not exploit knowledge of the contact process. Then, among all contact-probing strategies with the same average contact-probing interval, the strategy that probes at constant intervals achieves the minimum missing probability.

Proof: Consider a large interval of time L. Let us consider all strategies that probe the environment n times in this interval. As shown previously, for the strategy that probes at constant intervals $T = \frac{L}{n}$, the missing probability over duration L is $P_{\text{miss}}(T) = \frac{1}{T}\int_0^T F_D(x)dx$. Assume that an arbitrary scheme probes n times at $t_1, t_2, \ldots, t_n$. Define $t_0 = 0$ and $t_{n+1} = nT$. Then, we have $n+1$ intervals of $I_1 = t_1 - t_0, I_2 = t_2 - t_1, \ldots, I_{n+1} = t_{n+1} - t_n$. Since the scheme selects probe time t_i without knowledge of the contact process, the expected number of missed contacts in an interval of $[t_{i-1}, t_i]$ is $\lambda\int_0^{I_i} F_D(x)dx$, for $i \neq n+1$. All the contacts that occur in $[t_n, nT]$ will be missed. The expected missing probability is

$$\hat{P}_{\text{miss}} = \frac{1}{\lambda L}\left[\sum_{i=1}^{n} \lambda\int_0^{I_i} F_D(x)dx + \lambda I_{n+1}\right] \tag{3}$$

since the expected number of contacts arriving in duration L is λL. For $I_i \geq T$, we have

$$\begin{aligned}\lambda\int_0^{I_i} F_D(x)dx &= \lambda\left[\int_0^T F_D(x)dx + \int_T^{I_i} F_D(x)dx\right] \\ &\geq \lambda\int_0^T F_D(x)dx + \lambda\int_T^{I_i} F_D(T)dx \\ &= \lambda\int_0^T F_D(x)dx + \lambda(I_i - T)F_D(T). \end{aligned}\tag{4}$$

Similarly, when $I_i < T$ we have

$$\begin{aligned}\lambda\int_0^{I_i} F_D(x)dx &= \lambda\left[\int_0^T F_D(x)dx - \int_{I_i}^T F_D(x)dx\right] \\ &\geq \lambda\int_0^T F_D(x)dx - \lambda\int_{I_i}^T F_D(T)dx \\ &= \lambda\int_0^T F_D(x)dx + \lambda(I_i - T)F_D(T). \end{aligned}\tag{5}$$

We also have

$$\lambda I_{n+1} \geq \lambda I_{n+1}F_D(T). \tag{6}$$

Putting all these into (3), we have

$$\begin{aligned}\hat{P}_{\text{miss}} &\geq \frac{1}{\lambda nT}\left[\sum_{i=1}^{n}\lambda\left(\int_0^T F_D(x)dx + (I_i - T)F_D(T)\right)\right.\\ &\quad \left. + \lambda I_{n+1}F_D(T)\right]\\ &= \frac{1}{nT}\left[\sum_{i=1}^{n}\int_0^T F_D(x)dx + F_D(T)(\sum_{i=1}^{n+1} I_i - nT)\right]\\ &= P_{\text{miss}}(T). \end{aligned} \tag{7}$$

■

D. Tradeoffs in Energy Efficiency and P_{miss}

Having established that a constant contact-probing interval is optimal under certain assumptions, we now explore the tradeoff between energy efficiency and the missing probability. When contact durations are distributed according to a given distribution, we can analytically determine the relationship between T and $P_{\text{miss}}(T)$ according to (1). Here, we demonstrate the relationship between energy efficiency and missing probability for several typical distributions. In Section IV, we will focus on distributions obtained from real-world Bluetooth contact logs.

1) Exponential Distribution: When the contact duration is exponentially distributed, we have $F_D(x) = 1 - e^{-\mu x}$. Using (1), we have

$$P_{\text{miss}}(T) = \frac{\mu T - 1 + e^{-\mu T}}{\mu T}. \tag{8}$$

2) Uniform Distribution: The uniform distribution is

$$F_D(x) = \begin{cases} 0, & x < 0 \\ \frac{\mu x}{2}, & 0 \leq x \leq 2/\mu \\ 1, & x > 2/\mu \end{cases}. \tag{9}$$

Additionally, we have

$$P_{\text{miss}}(T) = \begin{cases} \frac{\mu T}{4}, & T < \frac{2}{\mu} \\ \frac{\mu T - 1}{\mu T}, & T \geq \frac{2}{\mu} \end{cases}. \tag{10}$$

3) Pareto Distribution: We have

$$F_D(x) = \begin{cases} 0, & x < \tau \\ 1 - (x/\tau)^{-k}, & x \geq \tau \end{cases}. \tag{11}$$

In this case, we have $1/\mu = k\tau/(k-1)$ when $k > 1$. The mean is unbounded when $k \leq 1$. Using (1), we have

$$P_{\text{miss}}(T) = 1 + \frac{k\tau}{T(1-k)} - \frac{\tau^k}{T^k(1-k)}, \quad T > \tau \tag{12}$$

Fig. 2 shows the tradeoff between energy consumption and missing probability for these distributions. The energy consumption is computed as $\frac{1}{\mu T}$, where we set $e_p = 1$ and normalize the energy consumption rate by the average contact duration of μ. We see that for exponential and uniform distributions, the missing probability of 5%–10% is near the knee of the curve that is a good tradeoff point between energy consumption and missing probability. This means the contact-probing interval should be around 1/6 to 1/3 of the average contact duration. However, for Pareto distribution, the contact-probing interval should be around τ to achieve a near-zero missing probability. In other words, for a constant arrival rate and a Pareto contact duration distribution, it is difficult to tradeoff between P_{miss} and T.

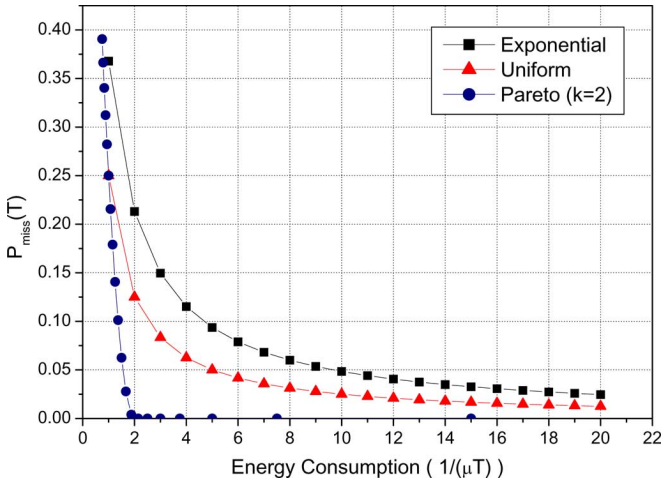


Fig. 2. Tradeoff between energy consumption and missing probability on different distributions.

E. Double Detection

As we stated earlier, a contact between device *A* and *B* is missed only if neither device probes the environment during the contact. Consider the case when two users *A* and *B* are independently probing the environment. Assume that both users are using the same constant contact-probing interval of T. Then, the probability that *A* does not discover *B* is $P_{\text{miss}}(T)$. However, the probability that neither *A* nor *B* discovers the other during a contact is much higher than $P^2_{\text{miss}}(T)$, even though their contact-probing processes are independent. Suppose one user probes at times of $T, 2T, \ldots, nT$, and the other probes at $y, y+T, \ldots, y+(n-1)T$. Without loss of generality, we can assume that $y < T/2$. Then, the probability that during a contact, neither user discovers the other is given by

$$\tilde{P}_{\text{miss}}(T, y) = \frac{1}{T}\left[\int_0^y F_D(x)dx + \int_0^{T-y} F_D(x)dx\right].$$

When $y = T/2$, $\tilde{P}_{\text{miss}}(T, y)$ has a minimum value of $P_{\text{miss}}(T/2)$, and $\tilde{P}_{\text{miss}}(T, y)$ has maximum value of $P_{\text{miss}}(T)$ when $y = 0$. Since the two users are probing independently, y is uniformly distributed in $[0, T/2]$, and the average missing probability is

$$\begin{aligned}&\tilde{P}_{\text{miss}}(T)\\ &= \frac{2}{T}\int_0^{\frac{T}{2}} \frac{1}{T}\left[\int_0^y F_D(x)dx + \int_0^{T-y} F_D(x)dx\right]dy\\ &= \frac{2}{T^2}\left[\int_0^{\frac{T}{2}}\int_0^y F_D(x)dxdy + \int_0^{\frac{T}{2}}\int_0^{T-y} F_D(x)dxdy\right]\\ &= \frac{2}{T^2}\left[\int_0^{\frac{T}{2}}\int_0^y F_D(x)dxdy + \int_{\frac{T}{2}}^{T}\int_0^{y} F_D(x)dxdy\right]\\ &= \frac{2}{T^2}\int_0^T\int_0^y F_D(x)dxdy. \end{aligned} \tag{13}$$

For example, when the contact duration is uniformly distributed as shown in (9), we have $\tilde{P}_{\text{miss}}(T) = \frac{\mu T}{6} = \frac{2}{3}P_{\text{miss}}(T)$, which is smaller than the single user missing probability only by a constant factor.

III. Nonstationary InterContact Times

Until now, we have made the simplifying assumption that the intercontact times are stationary. This assumption is clearly not true. For example, one would expect the intercontact time patterns late at night to differ significantly from the intercontact time distribution during lunch hours or peak traffic hours. This implies that the optimal contact-probing interval will vary with time. In the following sections, we study the contact arrival processes where the average contact arrival rate varies with time. In this case, we assume that the contact arrival rate is constant over some short time period of duration L. As before, we assume the contact durations are *i.i.d.*. This assumption will be justified in Section VI.

A. Problem Formulation

We divide the time to slots with equal duration of L. We denote the average contact arrival rate in the ith time slot as λ_i.

We proved in Theorem 1 that using a constant contact-probing interval of T_i is optimal during the ith time slot. In this case, the expected number of contacts captured during time period i will be $\lambda_i(1 - P_{\text{miss}}(T_i))$. However, to minimize the overall contact-missing probability, the optimal probing intervals T_i could be different for different time slots that have a different arrival rate of λ_i. Intuitively, more energy should be spent in time slots with high arrival rates to capture more contacts so that the overall contact missing probability can be minimized.

Consider the problem of maximizing the number of contacts captured over M time slots, given a constraint on the total energy used in the M time slots. This problem is equivalent to minimizing the contact-missing probability over the M time slots. As each probe consumes equal energy, the constraint on total energy consumption can be converted to a constraint of total number of probes used in the M time slots. Assume that the device can take N probes during the M time slots. We can then formulate the following optimization problem to find the optimal probing interval T_i for each time slot:

$$\begin{aligned} \text{Maximize} \quad & \sum_{i=1}^{M} \lambda_i(1 - P_{\text{miss}}(T_i)) \\ \text{s.t.} \quad & \sum_{i=1}^{M} \frac{L}{T_i} \le N \qquad T_i \ge 0 \quad \forall i. \end{aligned} \tag{14}$$

Defining variable x_i as $1/T_i$ and $U_i(x_i) = \lambda_i(1-P_{\text{miss}}(1/x_i))$, the optimization in (14) can be recast as

$$\begin{aligned} \text{Maximize} \quad & \sum_{i=1}^{M} U_i(x_i) \\ \text{s.t.} \quad & \sum_{i=1}^{M} Lx_i \le N \qquad x_i \ge 0 \quad \forall i. \end{aligned} \tag{15}$$

It can be verified that $U(x_i)$ is always a concave and increasing function for the three distributions we discussed above. Using the Karush–Kuhn–Tucker (KKT) conditions [16], we have

$$U_i'(x_i^*) = c \tag{16}$$

for all nonzero optimal solutions x_i^*. Note that $x_i^* = 0$ when $U_i'(0) \le c$. The energy cost c is a constant that can be obtained by solving the dual of (15). The optimal T_i^* can then be solved from

$$T_i^{*2} P_{\text{miss}}'(T_i^*) = \frac{c}{\lambda_i} \tag{17}$$

for a given λ_i. In practice, we can set the parameter c instead of N to achieve a certain tradeoff point between energy and missing probability. A larger value of c means energy is more precious (smaller N) and less energy will be used in the M time slots. Consequently, the overall contact-missing probability will be larger. Once the parameter c is set, we can use (17) to calculate the optimal T_i^* based on the estimate of λ_i for a time slot i. This solution minimizes the missing probability over all time, subject to energy constraints, which are determined by the energy cost c.

Example 1: Let the contact duration distribution be uniform, $L = 1$ h, and N is the number of times the device can probe in 24 h. Then, the optimal contact-probing interval T_i^* for the arrival rate λ_i is

$$T_i^* = \sqrt{\frac{4c}{\mu\lambda_i}} \tag{18}$$

by (17). If the λ_i are all equal, then clearly $T_i^* = \frac{24}{N}, \forall i$. Therefore, $c = \frac{\mu\lambda}{4}(\frac{24}{N})^2$ and $P_{\text{miss}}^*(T_i^*) = \frac{6\mu}{N}$ when $T_i^* \le \frac{2}{\mu}$.

B. Discussion

In Section II, we show that a constant probing rate is optimal for stationary processes, and the optimal tradeoff points can be derived from (1). In this section, we show that the average missing probability can be further reduced when the contact arrival rate is time-varying. When the contact arrival rates λ_i are different in different time slots, we can use *adaptive contact probing* to redistribute the limited probing energy according to the contact arrival rate. For a time slot with high contact arrival rate, we can increase the probing rate and reduce the missing probability so that most contacts arriving in this time slot can be captured. When the contact arrival rate is low, the contact-probing rate can be reduced since the number of missed contacts $\lambda_i L P_{\text{miss}}(T_i)$ will be small for a small λ_i, even if P_{miss} is quite high. In this way, the overall missing probability $\frac{\sum_{i=1}^{M} \lambda_i L P_{\text{miss}}(T_i)}{\sum_{i=1}^{M} \lambda_i L}$ can be minimized. However, the missing probability for time slots with low arrival rates could be very high in this case.

We now make two observations regarding the the optimization in (14). First, when the contact arrival rates are equal for all time slots, it is not necessary to use different probing intervals in different time slots, and a constant probing rate is optimal. In this scenario, the solution to (14) reduces to the solution for

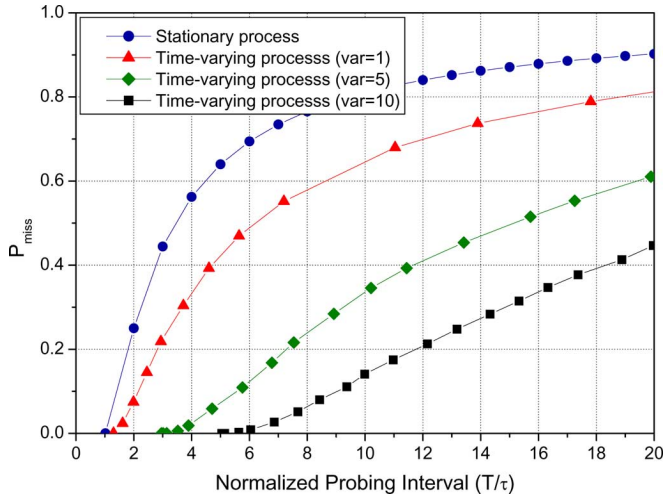


Fig. 3. Optimal missing probability for time-varying arrival process.

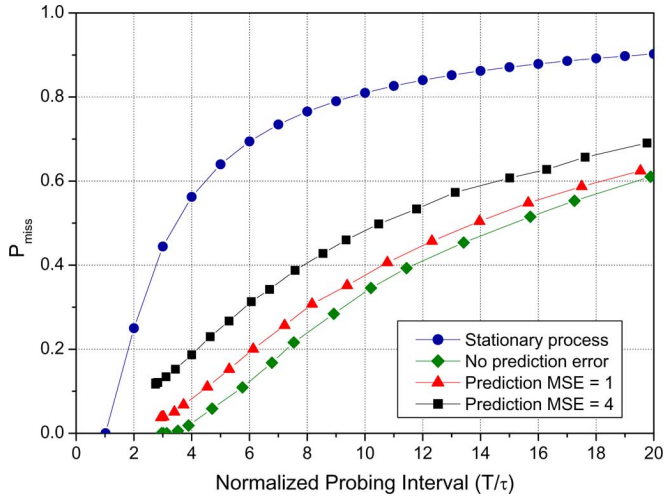


Fig. 4. Missing probability for time-varying arrival process (var $= 5$) with different prediction mean squared errors.

stationary intercontact times, derived in Section II. This is illustrated in Example 1.

Second, with time-varying contact arrival rates, the optimization algorithm should have enough information to predict the arrival rate λ_i. For a fixed value of c, the algorithm needs the arrival rate in the next time slot to calculate the optimal probing interval from (17). If the algorithm does not correctly predict the contact arrival rate λ_i in the next slot, the probing rate calculated based on the predicted λ_i will not be optimal.

We use an example to demonstrate the performance of adaptive contact probing in time-varying processes. Fig. 3 shows the performance for time-varying processes with different variances in contact arrival rates λ_i. In this example, the contact duration is distributed as a Pareto distribution with $k = 2$. For contact arrival rates λ_i, we assume they are random variables distributed according to a truncated Gaussian distribution.[3] We vary the mean and variance of the Gaussian distribution to change the variance of the contact arrival rate while keeping the mean arrival rate to 1. To expand the curve, we change the x-axis to normalized probing interval T/τ instead of the energy consumption rate $1/(\mu T)$ as in Fig. 2. In Fig. 3, a smaller value of P_{miss} for the same normalized probing interval represents better performance. Compared to the constant probing interval scheme used for stationary processes, we see that adaptive contact-probing scheme performs better when the variance of contact arrival rate increases. For a missing probability of 20%, the average contact-probing interval for processes with arrival rate variance of 10 is about 12, which is nearly six times longer than the probing interval used by the constant probing interval scheme. As we assume each probe uses the same amount of energy, the adaptive contact-probing scheme uses only 1/6 energy compared to the constant probing interval scheme.

In Fig. 3, we assume that there are no prediction errors in λ_i. Fig. 4 demonstrates the performance of adaptive probing scheme under different prediction mean squared errors (MSEs). We see that when the prediction error increases, the performance becomes worse. However, the performance is still much better than the constant probing interval scheme, even for large prediction errors. This suggests that we can use the adaptive contact-probing algorithm even when we cannot accurately predict the arrival rate.

[3]Note that the arrival rate cannot be negative; all negative arrival rates are set to zero.

IV. Human Contact Processes

In Sections II and III, we derived analytical models for general contact processes, which can be applied to different scenarios, such as contacts between humans or between mobile robots. In the next three sections, we mainly focus on the human contact process based on a set of Bluetooth contact logging experiments.

A. Data Collection Experiments

In order to characterize contact distributions, we handed Bluetooth phones to nine volunteers. We also installed static Bluetooth probes in high-traffic areas. The phones had a J2ME program running on them that initiated a Bluetooth device discovery every 30 s. If other Bluetooth devices are discovered in the vicinity, then the time of contact and Bluetooth address were captured. Since the probing software could terminate due to lack of energy, crashes, etc., we captured the start time and end time of each probing session. This allowed us to capture contact duration and intercontact time distributions accurately. Overall, we did 424 man days of data collection, and 12 332 unique devices were discovered in our experiment. To the best of our knowledge, this is the largest volume of unique devices discovered compared to other comparable studies.[4] For the details of the experiment, please refer to [11].

For the nine volunteers, we have the complete contact log over the experiment period. Our analysis and simulations are based on the contact patterns of these nine volunteers. Each contiguous set of scans during which a device is discovered is counted as a *contact*. Assume that a device (say D) is discovered in m contiguous scans. Then, the duration of the contact with D is the difference between the time of D's discovery in the mth

[4]In the Haggle [14] studies, a maximum of 41 volunteers probing at 120-s intervals discovered 200 unique devices over five days. In the Serendipity [13] study, 100 volunteers probing at 5-min intervals discovered 2798 unique devices over a nine-month period.

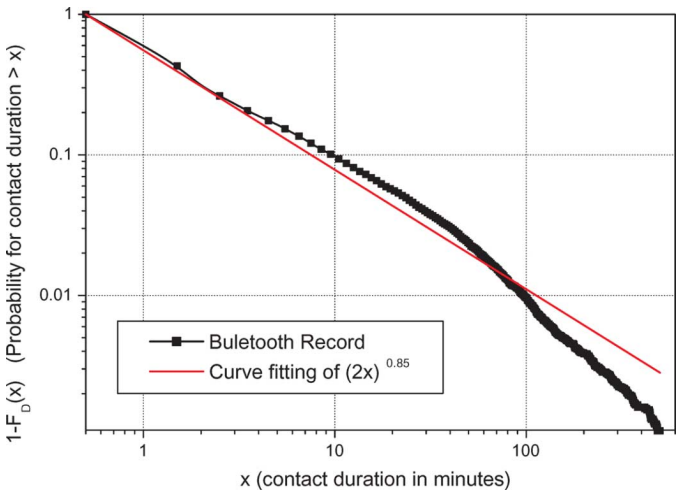


Fig. 5. Contact duration distribution.

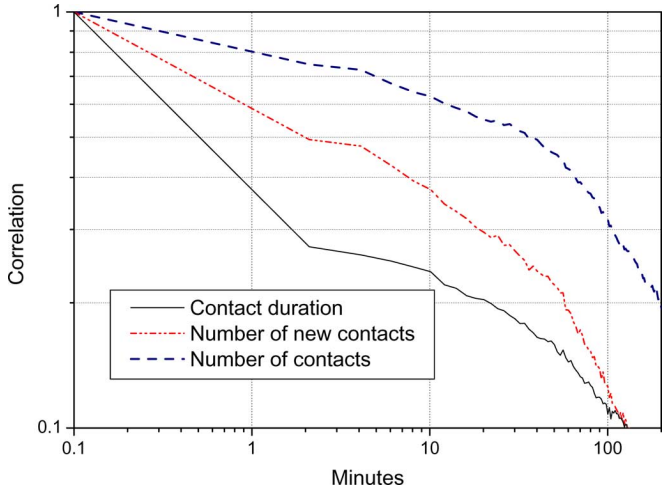


Fig. 6. Normalized correlation of contact duration and contact number over short time period ($\tau' = 2$ min).

scan and the first scan. If a device is only detected in one scan, we treat the duration of contact as 30 s.

B. Distribution of Contact Duration

We first look at contact duration distribution of the Bluetooth contact data. Fig. 5 plots the complementary cumulative distribution function (ccdf) curve, which is $1 - F_D(x)$ (probability of $\mathbf{t}_D > x$) in log-log scale. We see that the contact duration distribution follows a power law (Pareto distribution). By curve fitting, we can estimate $F_D(x) = 1 - (x/\tau)^{-k}$ with $\tau = 30$ s and $k = 0.85$. Although the mean for Pareto distribution with $k \leq 1$ is unbounded, the average contact duration is around 370 s in our case. This is because the probability of long contacts decreases rapidly when contact duration is longer than 2 h, and in our experiments, there are no contacts lasting longer than 5 h. The fact that the contact duration distribution follows a power law has also been verified by other studies (see [14] and references therein).

The fact that the contact duration distribution is Pareto-distributed is a pessimistic result. For the Pareto distribution, beyond a small threshold probing interval, the missing probability rises sharply from close to zero as shown in Fig. 2. Fortunately, as a consequence of human mobility, contact arrival rates are time-varying. As we show in later sections, this can be exploited to achieve significant energy savings.

C. Correlation Analysis

In order to see if there is any predictability in the contact patterns, which will aid us in designing good adaptive contact-probing mechanisms, we investigate the autocorrelation of the contact processes.

Here, we use the autocorrelation function [17]: Define an aggregated measurement series $f(i)$ as the average of certain measurements, over a time window of $[(i-1)\tau', i\tau']$. We calculate the autocorrelation of the measurement as $A_n = \sum_{i=1}^{\infty} f(i) \times f(i+n)$. In this paper, we consider the measurement of three types:

- *Contact duration*: The average contact duration for the contacts occurring in a time window.
- *Number of new contacts*: The number of newly arrived contacts during the time window of $[(i-1)\tau', i\tau']$. This measurement determines the correlation of intercontact time.
- *Number of contacts*: The number of contacts the device observes during the time window, including those that arrived in previous windows.

As shown in Fig. 6, the number of contacts and number of new contacts are highly correlated over short periods, and the correlation dropped faster when the correlation period is longer than 60 min. The contact duration is less correlated. The high correlation in the number of new contacts makes it possible to predict the contact arrival rate in future time slots, which will be discussed in Section V.

D. Correlation at Large Time Scales

The contact processes are highly related to human behavior. Therefore, we expect to see diurnal variations in the contact patterns. Also, since most people are creatures of habit, one might expect to see periodicity in the contact processes over daily or weekly time scales. Fig. 7 shows the correlation over long time periods. We clearly see the diurnal variations. The autocorrelation drops to zero over 12 h and had a period of 24 h. Finally, we do not see any significant correlation increase over a 168-h (one-week) period.

E. Self-Similar Nature

The slowly decreasing slope of the autocorrelation for the number of new contacts hints that the contact arrival process is self-similar [18], [19]. We test for self-similarity over four time scales of 10, 100, 1000, and 10 000 minutes. We use the R/S measurement to verify this conjecture. Let X_i denote the number of new contacts seen in 1 min and $Y_j = \sum_{i=1}^{j} X_i$ be the aggregated number of new contacts in j minutes. Define

$$R(t,L) = \max_{0 \leq i \leq L} \left[Y_{t+i} - Y_t - \frac{i}{L}(Y_{t+L} - Y_t) \right] - \min_{0 \leq i \leq L} \left[Y_{t+i} - Y_t - \frac{i}{L}(Y_{t+L} - Y_t) \right]$$

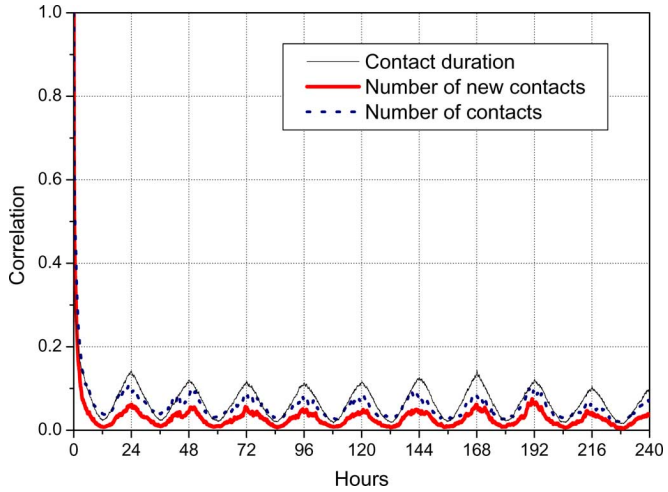


Fig. 7. Normalized correlation of contact duration and contact number over long time periods ($\tau' = 5$ min).

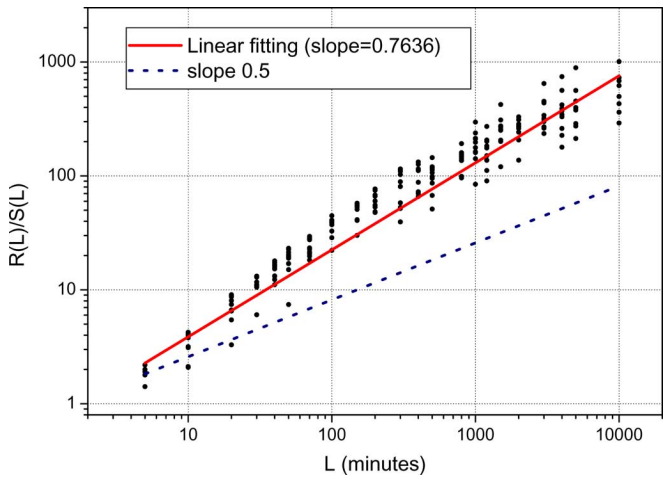


Fig. 8. The R/S statistic of the contact arrival process.

and

$$S(t,L) = \sqrt{\frac{1}{L}\sum_{i=t+1}^{t+L}(X_i - \overline{X_{t,L}})^2}, \overline{X_{t,L}} = \frac{\sum_{i=t+1}^{t+L} X_i}{L}.$$

If $\mathbb{E}\{R(L)/S(L)\} = CL^h$, with *Hurst* parameter $h \in (0.5, 1)$, then the process is self-similar [19]. From Fig. 8, we see that our process has a Hurst parameter close to 0.76.

This implies that the contact arrivals are not only long-range dependent but also bursty. An intuitive model for this is that arrivals follow an ON–OFF process where the ON and OFF durations have memory. This can potentially be exploited to achieve energy savings.

F. Mean and Variance of Contact Arrival Rate

As discussed in Section III, we can adapt the probing interval according to the contact arrival rate when the arrival rate is time-varying. Table I shows the mean and variance of the contact arrival rates at different time scales. Due to the self-similarity of the process, the variance increases faster than $CL^{1/2}$ when the time scale L increases. The normalized variance of the arrival rate is the variance divided by the mean arrival rate. As shown in Fig. 3, a larger normalized variance of the arrival rate leads to potentially better performance. In our human contact process, the normalized variance of the contact process is larger than 10 when the time slot size L is smaller than 1 h. Therefore, adaptive contact probing should perform better than the lowest curve in Fig. 3 when the arrival rate in the next hour is known.

TABLE I
CONTACT ARRIVAL RATE IN DIFFERENT TIME SCALES.

Time slot size	Mean arrival rate (contact/time slot)	Arrival rate variance	Normalized variance
5 minutes	0.2150	1.3350	28.8923
10 minutes	0.4299	4.4575	24.1172
30 minutes	1.2901	30.4295	18.2930
1 hour	2.5973	113.8747	17.1142

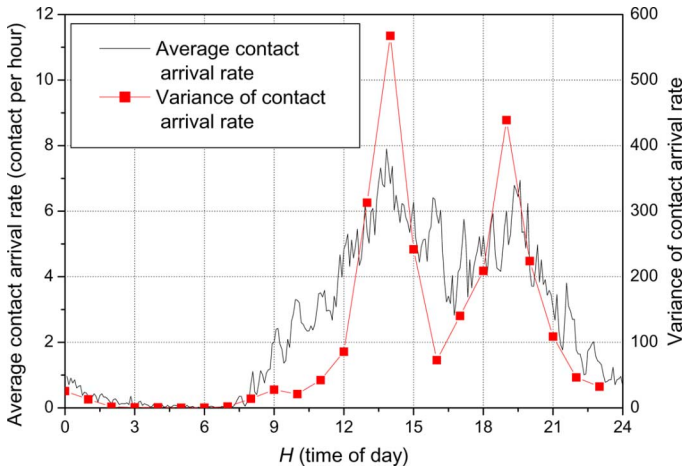


Fig. 9. Contact arrival rate and its variance over time of day.

G. Contact Arrival Rate versus Time of Day

The contact arrival rate distribution varies with time of day. In Fig. 9, we plot the average rate at which new contacts are seen at different times of the day. The contact arrival rate during the early morning is quite small. This implies that we can use longer contact-probing intervals during the early morning to save energy as described in Section III. The variance of the contact arrival rate over different days is also plotted in Fig. 9. We see that the variance is quite large. This shows the contact arrival rate at the same hour in different days may vary drastically. Although the arrival rates exhibit some patterns in a 24-h period, we cannot simply use the time of day to infer the arrival rate due to the large variance at the same time of day.

V. ADAPTIVE PROBING ALGORITHMS

In Section IV, we saw that human contact patterns have time-varying contact arrival rates when the time slot size L is smaller than 1 h. Also, the contact arrival rates have high short-term correlation and exhibit a 24-h periodic pattern. These observations suggest that the human contact process is both time-varying and predictable. We exploit these characteristics to design an efficient adaptive contact-probing scheme in this section.

Assumptions and Methodology

In this section, we investigate the design of adaptive contact-probing algorithms that can achieve a good tradeoff between the probability of a missed contact and energy consumption. As described in Section II-A, the energy consumption rate is inversely proportional to the average probing interval; e.g., a device with average probing interval of 30 s will consume two times more energy than a device with average probing interval of 60 s.

From our earlier analysis in Section III, we see that in order to compute the optimal contact-probing interval, we require k, the power law exponent for the contact duration distribution and the arrival rate. We contend that variations in the contact duration distribution, if any, will occur over large time scales. It is reasonable to hypothesize that contact duration distribution will always follow a power law but with different decay coefficients k over large time scales. This is validated by the fact that experiments at widely differing geographical locations have all displayed a power-law contact duration distribution but with different decay coefficients k [14], [12], [20]. For the rest of this section, we assume that the contact duration distribution is known and focus on the problem of adapting the contact-probing interval to variations in arrival rate. This assumption will be further justified in Section VI.

As shown in Section III, we can find the optimal value of the contact-probing interval by solving (17) when the arrival rate is known. Since our contact duration is Pareto-distributed, we have

$$T_i^* = \tau \left(\frac{c(1-k)}{\lambda_i k \tau} + 1 \right)^{\frac{1}{(1-k)}} \tag{19}$$

where $k = 0.85$ and $\tau = 30$ s for our logging data as shown in Section IV. With a different energy cost of c, we can achieve different tradeoff points between P_{miss} and T. Fig. 10 shows the optimal contact-probing interval for different arrival rates. We see that when the arrival rate is high, the optimal T_i^* quickly converges to τ. This means we can tolerate a certain amount of estimation error in λ_i when the contact arrival rate is high. However, when the contact rate is small, we need to estimate it accurately so that we will not underestimate the rate and use extremely long contact-probing intervals.

A. Contact Arrival Rate Estimation

Based on our analysis, we propose a class of adaptive contact-probing schemes based on short-term arrival rate estimation (STAR). STAR updates the estimate of the contact arrival rate λ_i periodically using a time slot length of L. The estimation is based on information available to the probing devices, such as time of day, statistics of historical arrival rates, and observed arrival rates in previous time slots. STAR then sets the probing interval for next probing according to the optimal T_i^* calculated from (19) using the estimated arrival rate $\hat{\lambda}_i$.

We will consider the following estimation methods for STAR and compare their performance in Section VI.

1) STAR-TOD: The STAR-time of day (STAR-TOD) scheme is motivated by two facts. First, people are creatures of habit. Second, Fig. 7 shows that the auto-correlation peaks at 24-h lags. These suggest that it may be possible to estimate the arrival rate using the time-of-day information.

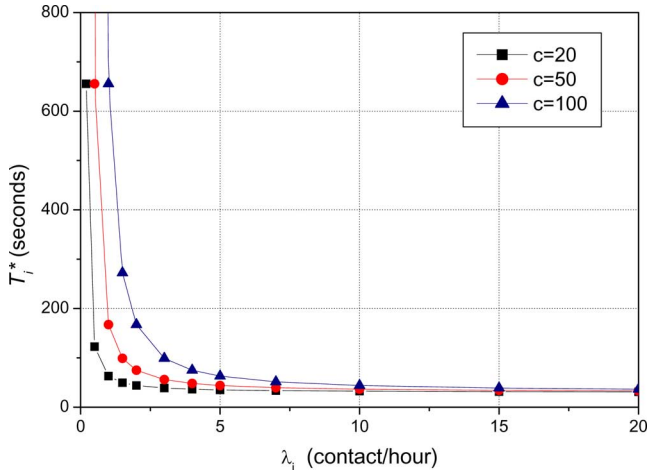


Fig. 10. Relationship between optimal contact-probing interval and the arrival rate.

STAR-TOD first collects the historical average arrival rate Λ_H at hour H (e.g., Λ_{12} is the arrival rate averaged over different days for the same time period of 1200–1259 h). Then, STAR-TOD reads the current hour H from the system clock of the device and sets

$$\hat{\lambda}_i = \Lambda_H. \tag{20}$$

In other words, STAR-TOD implicitly assumes the contact arrival rate is highly correlated to time of day, and the arrival rate will not vary dramatically in different days.

2) STAR-PTS: The STAR-previous time slot (STAR-PTS) scheme uses the detected arrival rate in the previous time slots as the estimation $\hat{\lambda}_i$. More specifically, it sets

$$\hat{\lambda}_i = \frac{\lambda_j}{i-j} \tag{21}$$

where λ_j is the nearest time slot with nonzero arrival rate. If the arrival rate in the last time slot λ_{i-1} is nonzero, (21) will set the estimate to $\hat{\lambda}_i = \lambda_{i-1}$.

Note that the historical arrival rate λ_j used in STAR-PTS comes from the arrival rate observed by the device. When T_j^* is large, the missing probability will be quite high, and λ_j could be inaccurate. To compensate this effect, we set λ_j as

$$\lambda_j = \frac{n}{L(1 - P_{\text{miss}}(T_j^*))} \tag{22}$$

where n is the number of new contacts detected by the device in time slot j.

STAR-PTS uses only the short-term history of the arrival process. This scheme may converge to certain nonoptimal operating points for human contact process. For example, after a night where no contacts arrive, the $\hat{\lambda}_i$ will converge to zero, and the T_i^* could be very long. In the next morning, it will be difficult to detect new contacts since the missing probability is high due to the large T_i^*. Therefore, STAR-PTS will keep on assuming λ_i is zero until it detects some contact by chance. To reduce this long "warm-up time" of STAR-PTS in the morning, we set $\hat{\lambda}_i$ as

$$\hat{\lambda}_i = w\frac{\lambda_j}{i-j} + (1-w)\Lambda_H \tag{23}$$

where w is the weight for different estimations. In (23), the time-of-day information is combined with estimation from short-term history to prevent the algorithm from getting stuck at the nonoptimal operating point. This approach combines both the long-term history of TOD and the short-term history. The long-term history can serve as a baseline of the everyday routine of the user. The short-term history will track dynamic changes in user behavior.

3) STAR-LMMSE: STAR linear minimum mean-squared error estimation (STAR-LMMSE) uses more short-term correlation information than STAR-PTS. It sets the estimation of the arrival rate in the next time slot to

$$\hat{\lambda}_i = \sum_{j=1}^{q} a_j \lambda_{i-j}. \tag{24}$$

Suppose the auto-correlation of arrival rates is

$$A_n = \sum_{i=1}^{\infty} \lambda_i \lambda_{i+n}. \tag{25}$$

We define

$$\mathbf{A} = \begin{bmatrix} A_0 & A_1 & A_2 & \dots & A_{q-1} \\ A_1 & A_0 & A_1 & \dots & A_{q-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ A_{q-1} & \dots & \dots & \dots & A_0 \end{bmatrix} \tag{26}$$

and

$$\mathbf{V} = [A_1\, A_2\, \dots\, A_q]. \tag{27}$$

Then setting

$$[a_1\, a_2, \dots, a_q] = V\left(\mathbf{A}^{-1}\right)^t \tag{28}$$

will minimize the MSE of the linear estimator [21], and the corresponding MSE is given by [21]

$$\mathbb{E}\{(\hat{\lambda}_i - \lambda_i)^2\} = A_0 - \sum_{j=1}^{q} a_j A_j. \tag{29}$$

Note that STAR-LMMSE only uses short-term history and may have the same "warm-up" problem as STAR-PTS. In this case, we can use the same weighted average method as discussed in STAR-PTS.

4) STAR-MMSE: We can also use the general nonlinear MMSE scheme to estimate the arrival rate by setting [21]

$$\hat{\lambda}_i = \mathbb{E}\{\lambda_i | H, \lambda_{i-1}, \dots, \lambda_{i-q}\}, \tag{30}$$

where H is the time of day as in STAR-TOD. In (30), H and λ_{i-j} can only take integer values. Therefore, the algorithm needs to store the $\hat{\lambda}_i$ corresponding to different combinations of H and λ_{i-j}s. As we can see from Fig. 10, the curve becomes flat when λ_i is large. Therefore, the algorithm can tolerate large estimation errors when λ_i is large. In practice, we can "quantize" large values of λ_{i-j}'s so that the number of combinations of H and λ_{i-j}'s can be reduced.

The STAR-MMSE algorithm uses both the short-term history and time-of-day information. However, it requires a large amount of memory to store the $\hat{\lambda}_i$ corresponding to different combinations of H and λ_{i-j}'s. Also, it needs a large amount of training data for calculating the conditional mean. Therefore, it has a much higher complexity compared to STAR-LMMSE and STAR-PTS. Though it may not be optimal in terms of minimizing missing probability, STAR-MMSE can serve as a performance benchmark for estimation-based contact-probing algorithms.

5) STAR-Genie: STAR-Genie is a noncausal algorithm that assumes that we know the actual arrival rate in the next time slot. It serves as a lower bound for the missing probability of all STAR-based algorithms.

VI. Simulations

A. Simulation Models

We now compare the performance of different adaptive contact-probing algorithms by running trace-driven simulations on the data that we have collected.

When calculating the missing probability, we use the Bluetooth contact logs, which are based on a 30-s contact-probing interval as the baseline. We apply different adaptive contact-probing algorithms to filter the contact log data. Assume that a device D has been logged in our experiments from time t to $t+\tau$. Then, for a specific adaptive contact-probing algorithm, we say that the contact has been missed if the algorithm does not initiate any probe in the interval $[t, t+\tau]$. The contact missing probability is computed as the ratio of the number of contacts missed by the adaptive contact-probing algorithm to the total number of contacts made in our data logging experiments.

B. Verification of Analysis

First, we verify that the relationship between T and P_{miss} derived in (12) is correct for a constant contact-probing interval T. From our logging data, we have computed the decay exponent of the contact distribution $k = 0.85$. Since the granularity of our logging experiments is over 30-s intervals, we set $\tau = 30$. The comparison between the trace-driven simulations and the analytical results are shown in Fig. 11. We see that the results of the simulation are quite close to the theoretical results. Since the constant contact-probing interval algorithm does not adapt to the changes in the contact arrival rate, we will use this as the lower bound for the average contact-probing interval to achieve a given missing probability.

C. Performance Comparison

We use data-driven simulations to compare the performance of different adaptive probing algorithms. In the simulation, we use a time slot size of 10 min based on the correlation of human contacts shown in Fig. 6. Since the correlation drops below 0.1 after 2 h, we use the arrival rate in the previous 2 h (i.e., $q = 12$) to estimate λ_i in STAR-LMMSE. For STAR-MMSE, we use $q = 4$ due to the large storage space and training data size required by the algorithm. Additionally, we assume STAR-MMSE has perfect knowledge of the arrival rates in the previous time slots. In this sense, it is a lower bound to the performance of the algorithm that uses the detected arrivals only.

Fig. 12 shows the relationship between missing probability and the long-term average probing interval for different algo-

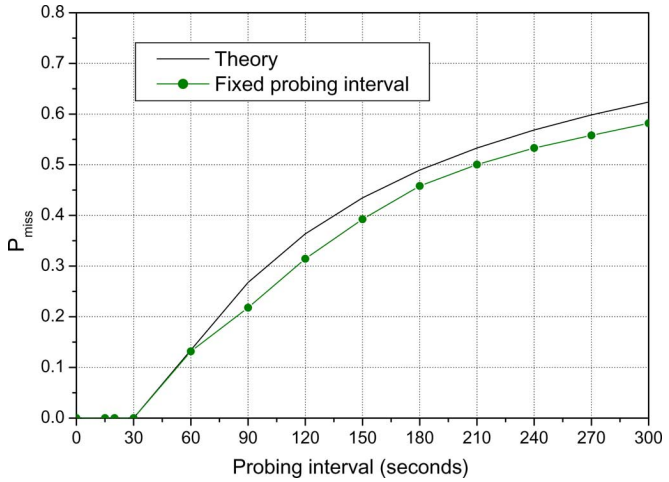


Fig. 11. Comparing the missing probability of the Bluetooth log to theory results.

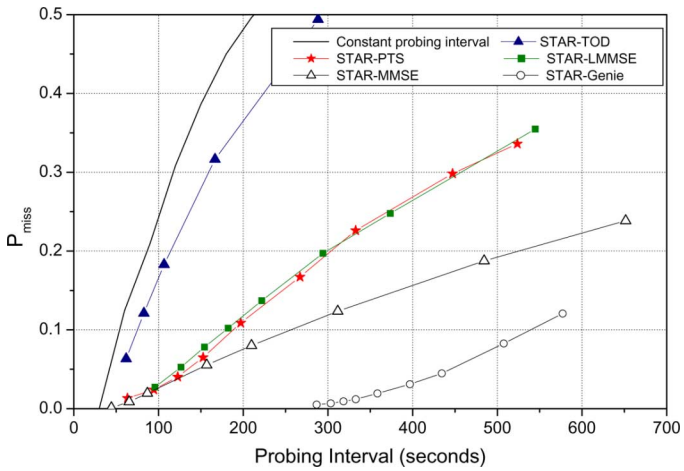


Fig. 12. Improving energy efficiency by adapting to the contact arrival rate.

rithms. The energy consumption rate of the algorithms can be calculated through e_p/T, where e_p is the energy used in each probe and T is the average probing interval.[5] The curve for STAR-Genie provides performance bounds for STAR-based adaptive contact-probing algorithms. STAR-Genie is a noncausal algorithm that has accurate future contact arrival rates. In other words, STAR-Genie shows the optimal tradeoff between energy and missing probability when there is no estimation error. Comparing the performance of STAR-Genie to the results in Fig. 3, we see that the missing probability of STAR-Genie is lower. This can be explained by noting that human contact processes have large variance ($\gg$ 10) in short-term contact arrival rates that can be exploited by STAR. Since STAR-Genie is noncausal, an implementable contact-probing algorithm needs to estimate the arrival rate.

STAR-TOD estimates the arrival rate based on the time-of-day information. The performance of STAR-TOD is slightly better than constant probing interval algorithms and not as good as other estimation methods that use short-term arrival history. This observation suggests that the time-of-day information is less important than short-term human mobility behavior in predicting contact arrival rates.

The performance of STAR-PTS and STAR-LMMSE is much better. For a missing probability of 20%, the probing interval for STAR-PTS and STAR-LMMSE is about 300 s, which is more than three times the probing interval of the constant probing interval algorithm. In other words, STAR-PTS and STAR-LMMSE use only 1/3 the energy for contact probing than constant probing interval schemes. Fig. 12 also shows that the performances of STAR-PTS and STAR-LMMSE are nearly the same. Compared to STAR-LMMSE, which estimates arrival rates based on history of the past 12 time slots, STAR-PTS only uses the nearest time slot in the estimation. The similar performances of STAR-LMMSE and STAR-PTS suggest that it may be good enough to use a lesser amount of history in the arrival rate estimation. We take advantage of this to reduce the complexity of STAR-MMSE.

The STAR-MMSE algorithm gives the best arrival rate estimation that can be achieved with perfect knowledge of historical arrival rates and the statistics of the contact patterns. The performance of STAR-MMSE is better than STAR-LMMSE when the average probing interval is large. For a missing probability of 20%, the probing interval of STAR-MMSE is over 1.5 times that of STAR-LMMSE and five times that of the constant probing scheme. However, the difference with STAR-LMMSE is negligible when the probing interval is small. This could have several explanations. First, STAR-LMMSE uses the detected arrival rate, while the ideal STAR-MMSE uses the true arrival rates in the estimation. Second, STAR-LMMSE only has the correlation of arrival rates, while STAR-MMSE has detailed information on the conditional distribution of the arrival rates. Third, STAR-LMMSE uses the time-of-day information in a heuristic weighted averaging manner, while STAR-MMSE uses the time-of-day information directly in the conditional mean estimator.

To further compare STAR-LMMSE and STAR-MMSE, we use the true value of λ_{i-j} in the estimator of STAR-LMMSE. The performance of STAR-LMMSE becomes better when the probing interval is large, as seen in Fig. 13. This is because the missing probability increases with the probing interval. When the missing probability is large, the detected arrival rate used in the estimation might be wrong. Consequently, the estimation error will increase when wrong historical arrival rates are used in estimation. Therefore, the performance of STAR-LMMSE can be greatly improved with accurate historical information when the probing interval is large. However, even with the actual historical arrival rates, the performance of STAR-LMMSE still cannot be as good as STAR-MMSE. This may due to the larger MSE of STAR-LMMSE. In our simulations, the MSE of STAR-LMMSE with accurate history is 2.2943, while the MSE of STAR-MMSE is 1.1280. Although STAR-MMSE gives the best performance, it requires much more storage and training data than STAR-LMMSE and STAR-PTS. Therefore, STAR-LMMSE and STAR-PTS may be more appropriate for use in mobile devices due to their simplicity.

D. Contact Duration Distribution

In our analysis in Section III, we assumed that the contact duration is distributed according to the Pareto distributions with different k. In real contact processes, the probing algorithm

[5]In our Bluetooth experiments, the value of e_p is around 2.1 J as the Bluetooth device-discovery process takes 10–20 s.

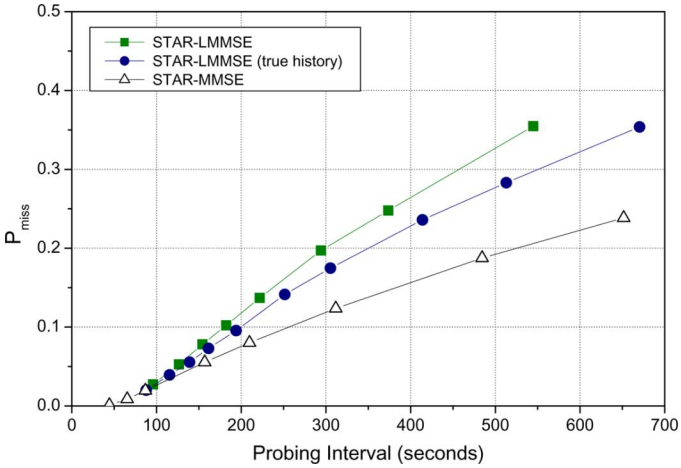


Fig. 13. STAR-LMMSE with true historical arrival rate.

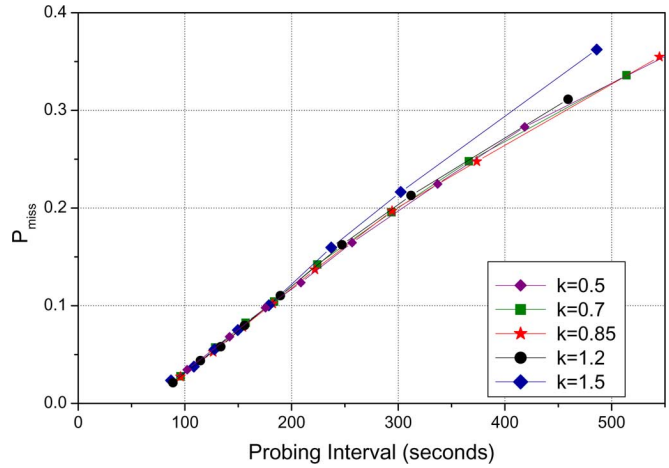


Fig. 14. STAR-LMMSE with different k values.

may not have the accurate value of k, which may also be time-varying. We use different k values in the STAR-LMMSE to see whether it is sensitive to changes in the value of k. Fig. 14 shows the performance of STAR-LMMSE running with k changing from 0.5 to 1.5. Although our data has a k value of 0.846, changing the value of k used in the algorithm does not deteriorate the performance much. This shows that the algorithm can work well with inaccurate or varying values of k.

VII. Related Work

User mobility has a profound impact on the performance of wireless networks. It has been shown that the capacity of wireless networks can be improved by random [4] or controlled [22] mobility. The impact of user mobility in delay performance and forwarding algorithm design has also been widely studied in DTNs that use Bluetooth or WLAN [5], [20], [23], [24]. In this paper, we study how user mobility will effect energy efficiency in device discovery for delay-tolerant applications. Our protocol provides an energy-conscious device-discovery service, which can be used by many delay-tolerant applications.

In [25], the authors observe that existing mobility models are too simple and do not accurately reflect user mobility. Based on WLAN traces from various universities, they observe that people have location preferences, which they visit periodically. They then propose an empirical model to simulate user mobility and validate it with WLAN data traces. The goal of this paper is modeling user mobility and contact patterns and not to derive contact-probing algorithms.

In [26], the authors again consider WLAN traces from the University of California, San Diego, and Dartmouth College, Hanover, MA. They analyze the statistics of interpair contact time distribution, i.e., the time between subsequent meetings of a specific pair of nodes. This metric is especially useful for DTNs. They show that the interpair contact time distribution is self-similar. In our work, we consider traces of a different process, namely the set of contacts made as a user moves around. This contact process is significantly different from that obtained from WLAN traces. In WLAN traces, contacts between users can be inferred only if two users associate with the same access point at the same time. Contacts between users at locations where there are no access points cannot be inferred. Also, we characterize the intercontact time, i.e., the time between the discovery of two new contacts. This metric is useful for both DTN and DTD applications.

Energy-efficient Bluetooth device discovery has been studied in [27] and [28]. However, they focus on the Bluetooth protocol stack. Our algorithms are independent of the communication protocol and adapt to user mobility to conserve energy for device discovery, making them applicable to many communication systems.

Stochastic-event capturing schemes are studied in wireless sensor networks [29]. In [29], optimal visiting routes are devised for mobile sensors to capture events that randomly happen in different points of interest. The event process is assumed to be memoryless, and the parameters of the process are known in advance. In our study, we investigate real-world contact processes that are more complicated than memoryless arrival and departure processes. The analysis in this paper can be applied to a general renewal process, and our algorithm can dynamically adapt to unknown parameters in the process.

VIII. Reflections

In this paper, we have identified that contact-probing mechanisms play a critical role in certain mobile delay-tolerant applications. In these applications, mobile devices periodically probe their environment for the presence of new contacts. We investigated the design of energy-conscious, adaptive contact-probing algorithms that trade off energy consumption and the probability of missing a contact. Our key contributions were: 1) a theoretical foundation that aids in the design of adaptive contact-probing algorithms; 2) real-world experiments and characterization of empirical contact patterns; and 3) design and validation of an adaptive probing algorithm (called STAR) via trace-driven simulations. We demonstrate that STAR-PTS is *three* times more energy-efficient than a naive constant-probing algorithm. We also show STAR-MMSE, an ideal MMSE estimator-based algorithm, could further improve the energy efficiency by 50% compared to STAR-PTS. We now reflect on what we have done.

A. Exploiting Contact Bursts

Our empirical data show that the contact duration is Pareto-distributed and the new contact arrivals are self-similar, meaning they are bursty. Not surprisingly, it is advantageous to exploit

the bursty nature to find a good tradeoff between missing probability and energy. This is what STAR implicitly does.

B. Other Heuristics

It is also possible to use other heuristics in adaptive contact-probing algorithms. For example, we can use the additive increase/multiplicative decrease (AIMD) algorithm in the adaptive contact-probing scheme. The AIMD algorithm additively increases the probing interval when no new contacts are discovered and decreases the probing interval by half when new contacts are detected. Although AIMD changes the probing interval according to the arrival rate as in STAR algorithm, the performance of AIMD is not as good as STAR-PTS, as shown in [1].

C. Unknown Contact Duration Distribution

In contact processes other than human contacts, the intercontact times could be stationary while the contact duration distribution could change. In such processes, it is hard to calculate P_{miss} from the observations of the device since we do not have the information on contact duration distribution. It is shown in [1] that we can use the percentage of short contacts, which are contacts that are only detected by the device in one probing and leave in the next probing, to bound the missing probability and adapt to the unknown contact duration distribution.

D. Maintaining Short-Term P_{miss}

Even though STAR maintains a low missing probability over large time scales, it can miss a relatively large fraction of contacts in the short term. This is because we have so far assumed that when a device is not probing, it has no way of knowing if new contacts have arrived even if those new contacts are probing. However, if devices are cognizant of the fact that they have been probed, this information can clearly be used to better adapt the probing interval and thereby reduce the short term contact missing probability. We point out that our analysis and simulations were motivated by current Bluetooth implementations, which do not expose the fact that a device has been probed, to the application. However, this constraint is not fundamental and can easily be taken into account.

E. Synthetic Contact Models

Our empirical data set for contact patterns is much larger than any comparable data set. We have learned quite a bit about mobile device contact patterns and are certain that there is much more that can be gleaned. In addition, we believe that our data set and analysis can be used to develop synthetic models for real-world contact patterns, as done in [30] (for modeling Internet traffic) and [25] (for modeling association with WiFi access points).

F. Energy Consumption During Contact Process

We only consider energy used for sending out probing messages in this paper. Energy used in other parts of the contact process should also be considered when designing the contact protocols. First, a device needs to periodically listen to the channel in order to be probed. Increasing the listening frequency can reduce the time duration of device discovery. Thus, energy used in each probe for the probing device can be reduced. However, the device to be probed will consume considerable energy in periodic listening. Therefore, there is a tradeoff between the listening frequency and energy consumption of sending probing messages [31]. Second, the energy used in information exchange after the contact has been detected should also be considered. In this case, the device may need to use additional information—such as the importance of the contact, expected amount of information exchange, and residual energy—to decide whether to take the contact or not.

Appendix
Contact Missing Probability When Probes are Unreliable

Consider the case that a probe can fail to detect a present contact with probability κ. We also assume that the failure probabilities of consequent probes are independent. For contacts with $\mathbf{t}_D \in [0,T]$, they will either not be probed at all or only be probed once during their contact duration. As shown in Section II, the probability that these contacts will not be probed is $\frac{T-\mathbb{E}\{\mathbf{t}_D|\mathbf{t}_D<T\}}{T}$. Therefore, the probability that these contacts can be probed once is $\frac{\mathbb{E}\{\mathbf{t}_D|\mathbf{t}_D<T\}}{T}$. As probes can fail with probability of κ, the missing probability given that $\mathbf{t}_D \in [0,T]$ is

$$P_{\text{miss}}(T|0 \le \mathbf{t}_D < T) = \frac{T-\mathbb{E}\{\mathbf{t}_D|\mathbf{t}_D<T\}}{T} + \kappa\frac{\mathbb{E}\{\mathbf{t}_D|\mathbf{t}_D<T\}}{T}. \quad (31)$$

Similarly, for contacts with $\mathbf{t}_D \in [nT,(n+1)T]$, they can be probed by n times or $n+1$ times during their contact duration. If a contact has been probed for n times, the probability of missing that contact is κ^n for independent probes. Define $E_n = \mathbb{E}\{\mathbf{t}_D|nT \le \mathbf{t}_D < (n+1)T\} - nT$. We have

$$P_{\text{miss}}(T|nT \le \mathbf{t}_D < (n+1)T) = \kappa^n\frac{T-E_n}{T} + \kappa^{n+1}\frac{E_n}{T}. \quad (32)$$

Therefore

$$\begin{aligned} &P_{\text{miss}}(T) \\ &= \sum_{n=0}^{\infty} P_{\text{miss}}(T|nT \le \mathbf{t}_D < (n+1)T)(F_D((n+1)T) - F_D(nT)) \\ &= \sum_{n=0}^{\infty} \kappa^n \left(\frac{T-E_n}{T} + \kappa\frac{E_n}{T}\right)(F_D((n+1)T) - F_D(nT)). \end{aligned}$$

Note that P_{miss} only depends on $F_D(x)$ and κ in this case.

Acknowledgment

The authors wish to thank F. Kohwaja for the implementation on mobile phones.

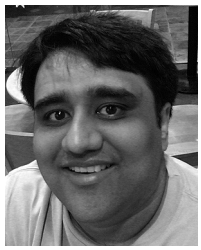
References

[1] W. Wang, V. Srinivasan, and M. Motani, "Adaptive contact probing mechanisms for delay tolerant applications," in *Proc. ACM MobiCom*, 2007, pp. 230–241.
[2] K. Fall, "A delay-tolerant network architecture for challenged internets," in *Proc. ACM SIGCOMM*, 2003, pp. 27–34.
[3] S. Jain, K. Fall, and R. Patra, "Routing in a delay tolerant network," *SIGCOMM Comput. Commun. Rev.*, vol. 34, no. 4, pp. 145–158, 2004.
[4] M. Grossglauser and D. Tse, "Mobility increases the capacity of ad hoc wireless networks," *IEEE/ACM Trans. Netw.*, vol. 10, no. 4, pp. 477–486, Aug. 2002.

[5] 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, pp. 1–13.
[6] M. Motani, V. Srinivasan, and P. Nuggenhalli, "Peoplenet: Engineering a wireless virtual social network," in *Proc. ACM MobiCom*, 2005, pp. 243–257.
[7] A. J. Campbell, S. B. Eisenman, N. D. Lane, E. Miluzzo, and R. A. Peterson, "People-centric urban sensing," in *Proc. WICON*, 2006, vol. 220, Article No. 18.
[8] H. Y. O. Wolfson, B. Xu, and H. Cao, "Search-and-discover in mobile p2p network databases," in *Proc. IEEE ICDCS*, 2006, pp. 65–73.
[9] M. Kuhn and R. Wattenhofer, "Community aware mobile networking," in *Proc. 1st Workshop on Mobile Services and Personalized Environ.*, 2006, pp. 1–14.
[10] V. Srinivasan, M. Motani, and W. T. Ooi, "Analysis and implications of student contact patterns derived from campus schedules," in *Proc. ACM MobiCom*, 2006, pp. 86–97.
[11] A. Natarajan, M. Motani, and V. Srinivasan, "Understanding urban interactions from Bbluetooth phone contact traces," in *Proc. PAM*, 2007, pp. 115–124.
[12] J. Su, A. Chin, A. Popivanova, and E. d. L. A. Goel, "User mobility for opportunistic ad-hoc networking," in *Proc. IEEE WMCSA*, 2004, pp. 41–50.
[13] N. Eagle and A. Pentland, "Social serendipity: Mobilizing social software," *IEEE Pervasive Comput.*, vol. 4, no. 2, pp. 28–34, 2005.
[14] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Pocket switched networks: Real world mobility and its consequences for opportunistic forwarding," Univ. Cambridge, Cambridge, U.K., 2005, Tech. Rep. No. 617.
[15] S. M. Ross, *Stochastic Processes*. New York: Wiley, 1983.
[16] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[17] G. He and J. Hou, "An in-depth, analytical study of sampling techniques for self-similar internet traffic," in *Proc. IEEE ICDCS*, 2005, pp. 404–413.
[18] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of ethernet traffic," *IEEE/ACM Trans. Netw.*, vol. 2, no. 1, pp. 1–15, Feb. 1994.
[19] J. Beran, *Statistics for Long-Memory Processes*. London, U.K.: Chapman & Hall, 1994.
[20] M. McNett and G. M. Voelker, "Access and mobility of wireless PDA users," *Mobile Comput. Commun. Rev.*, vol. 9, no. 2, pp. 40–55, 2005.
[21] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed. New York: McGraw Hill, 2002.
[22] R. Shah, S. Roy, S. Jain, and W. Brunette, "Data mules: Modeling a three-tier architecture for sparse sensor networks," in *Proc. IEEE SNPA*, 2003, pp. 30–41.
[23] 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 on Delay-Tolerant Netw.*, 2005, pp. 244–251.
[24] T. Gross and C. Tuduce, "A mobility model based on WLAN traces and its validation," in *Proc. IEEE INFOCOM*, 2005, pp. 664–674.
[25] W.-J. Hsu, T. Spyropoulos, K. Psounis, and A. Helmy, "Modeling time-variant user mobility in wireless mobile networks," in *Proc. IEEE INFOCOM*, 2007, pp. 758–766.
[26] L.-J. Chen, Y.-C. Chen, T. Sen, P. Sreedevi, K.-T. Chen, C.-H. Yu, and H.-H. Chu, "Finding self-similarities in opportunistic people networks," in *Proc. IEEE INFOCOM*, 2007, pp. 2286–2290.
[27] C. Drula, "Fast and energy efficient neighbour discovery for opportunistic networking with Bluetooth," Master's thesis, Univ. of Toronto, Toronto, ON, Canada, 2005.
[28] G. Zaruba and I. Chlamtac, "Accelerating Bluetooth inquiry for personal area networks," in *Proc. IEEE GLOBECOM*, 2003, pp. 702–706.
[29] N. Bisnik, A. Abouzeid, and V. Isler, "Stochastic event capture using mobile sensors subject to a quality metric," in *Proc. ACM MobiCom*, 2006, pp. 98–109.
[30] P. Barford and M. Crovella, "Generating representative web workloads for network and server performance evaluation," in *Proc. ACM Sigmetrics/Performance*, 1998, pp. 151–160.
[31] J. Polastre, J. Hill, and D. Culler, "Versatile low power media access for wireless sensor networks," in *Proc. ACM SenSys*, 2004, pp. 95–107.

Wei Wang (S'06–M'09) received the Ph.D. degree in electrical and computer engineering from the National University of Singapore in 2008.

He is currently with Microsoft Research Asia, Beijing, China. His research interests are in the area of wireless networking.



Mehul Motani (S'92–M'00) received the B.E. degree from Cooper Union, New York, the M.S. degree from Syracuse University, Syracuse, NY, and the Ph.D. degree from Cornell University, Ithaca, NY, all in electrical and computer engineering.

He is currently an Assistant Professor in the Electrical and Computer Engineering Department at the National University of Singapore. Previously, he was a Research Scientist at the Institute for Infocomm Research in Singapore for three years and a Systems Engineer at Lockheed Martin in Syracuse, NY, for over four years. Recently, he has been working on research problems that sit at the boundary of information theory, communications, and networking, including the design and analysis of wireless ad hoc and sensor network systems.

Dr. Motani was awarded the Intel Foundation Fellowship for work related to his Ph.D. in 2000. He has served on the organizing and technical program committees of ISIT, MobiCom, INFOCOM, SECON, and various other conferences. He participates actively in the IEEE and Association for Computing Machinery and has served as the Secretary of the IEEE Information Theory Society Board of Governors.

Vikram Srinivasan (M'03) received the Bachelor of Science degree in physics from the University of Chennai, Chennai, India, in 1994; the M.E. degree in electrical communications engineering from the Indian Institute of Science, Bangalore, India, in 1998; and the Ph.D. degree in electrical and computer engineering from the University of California at San Diego in 2003.

He was an Assistant Professor in the Electrical and Computer Engineering Department, National University of Singapore, from 2003 to 2007. He is currently with Bell Labs Research, Bangalore, India. His research interests are broadly in the area of wireless networks.