

Mining Frequent Trajectory Patterns for Activity Monitoring Using Radio Frequency Tag Arrays

Yunhao Liu, *Senior Member, IEEE*, Yiyang Zhao, *Member, IEEE*, Lei Chen, *Member, IEEE*, Jian Pei, *Senior Member, IEEE*, and Jinsong Han, *Member, IEEE*

Abstract—Activity monitoring, a crucial task in many applications, is often conducted expensively using video cameras. Effectively monitoring a large field by analyzing images from multiple cameras remains a challenging issue. Other approaches generally require the tracking objects to attach special devices, which are infeasible in many scenarios. To address the issue, we propose to use RF tag arrays for activity monitoring, where data mining techniques play a critical role. The RFID technology provides an economically attractive solution due to the low cost of RF tags and readers. Another novelty of this design is that the tracking objects do not need to be equipped with any RF transmitters or receivers. By developing a practical fault-tolerant method, we offset the noise of RF tag data and mine frequent trajectory patterns as models of regular activities. Our empirical study using real RFID systems and data sets verifies the feasibility and the effectiveness of this design.

Index Terms—Active RFID, mining, trajectory

1 INTRODUCTION

IN many applications, it is necessary to monitor activities in closed fields. For example, in chemical plants or large industrial workshops, security control staffs have to monitor “suspicious” activities. Oftentimes, in these applications, the monitoring area is very large and activities (moving trajectories) are sparse. Intuitively, the normal trajectories of moving objects often follow regular patterns. Once we have these patterns, abnormal behaviors of moving objects can be easily detected through pattern matching [1].

Currently, activity monitoring is widely conducted using video monitoring equipment such as digital cameras. Cameras are expensive while each camera can only cover a small area and specific trails. As illustrated in Fig. 1, a small part of the large surveillance area is monitored. In contrast, shadowed parts indicate the places without monitoring, from where unauthorized persons or objects may break through. Moreover, it is hard to automatically

analyze the activity patterns in a large field with images from multiple cameras.

Monitoring with video cameras has following limitations. First, the target trajectories must be predefined. Once the trajectories change, the cameras may need to be redeployed. Indeed, the frequent trajectories may not be known and they frequently change over time in many situations. Second, except for the target trajectories, monitoring other regions is difficult. Third, automatically analyzing the images from multiple cameras and detecting irregular activities is not trivial. And last, digital cameras are expensive. It is often a financial concern to deploy a large number of cameras.

We propose a novel application of the Radio Frequency Identification (RFID) technology to provide an inexpensive and relatively accurate approach to activity monitoring. By employing an array of RF tags and a few RF readers, we use data mining techniques to detect and analyze frequent trajectory patterns. We focus on extracting frequent patterns as these patterns can be used as domain knowledge to capture any anomalies.

Since RF tags and readers are much cheaper than cameras (in US dollars, an active RF tag is about 50 cents and an RF reader is several hundred dollars), and data mining techniques can detect frequent patterns online, our approach is more flexible and much cost-efficient than the video monitoring solutions.

1.1 RFID and Location Sensing

RFID is a means of storing and retrieving data through electromagnetic transmission to an RF compatible integrated circuit. It is now being seen as a radical means of enhancing data handling processes [2]. An RF reader can read data emitted from active RF tags. RF readers and tags use a defined radio frequency and protocol to transmit and receive data. RF tags are categorized as either passive or active. Passive RF tags operate without a battery. Their read ranges are very limited. Active tags contain both a radio

• Y. Liu is with the School of Software, Tsinghua National Lab for Information Science and Technology, Tsinghua University, and the Hong Kong University of Science and Technology. E-mail: Yunhao@GreenOrbs.com.

• Y. Zhao is with the Tsinghua National Laboratory for Information Science and Technology (TNLIST), School of Software, Tsinghua University, Beijing 100084, P.R. China. E-mail: yiyangzhao@tsinghua.edu.cn.

• L. Chen is with the Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Kowloon, Hong Kong. E-mail: leichen@cse.ust.hk.

• J. Pei is with the School of Computing Science, Simon Fraser University, 8888 University Drive, Burnaby, BC Canada V5A 1S6. E-mail: jpei@cs.sfu.ca.

• J. Han is with the School of Electronic and Information Engineering, Department of Computer Science and Technology, Institute of Computer Software and Theory, Xi'an Jiaotong University, No. 28 Xianning West Road, Xi'an, Shaanxi 710049, China. E-mail: hanjinsong@mail.xjtu.edu.cn.

Manuscript received 25 May 2011; revised 3 Oct. 2011; accepted 12 Oct. 2011; published online 14 Dec. 2011.

Recommended for acceptance by K. Li.

For information on obtaining reprints of this article, please send e-mail to: tpds@computer.org, and reference IEEECS Log Number TPDS-2011-05-0327. Digital Object Identifier no. 10.1109/TPDS.2011.307.

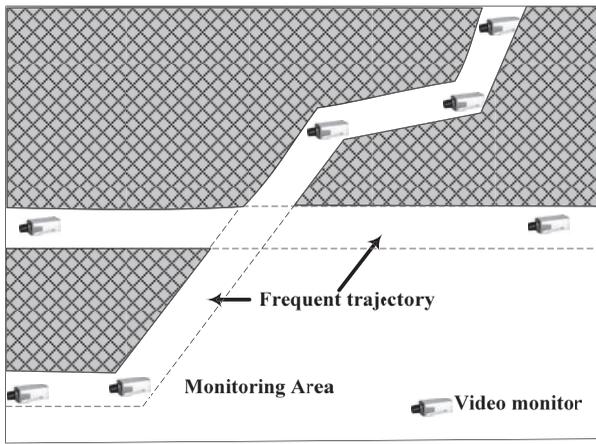


Fig. 1. Monitoring activities using video equipment.

transceiver and a button-cell battery to power the transceiver, and hence have larger ranges than passive tags.

We are interested in using commodity off-the-shelf products. There are several advantages of the RFID technology, including the no-contact and nonline-of-sight nature which is the common among all types of RFID systems [3]. All RF tags can be read despite extreme environmental factors such as snow, fog, ice, paint, and other challenging conditions [4].

The other advantages are their promising transmission ranges and cost-effectiveness. Indeed, if we deploy a video camera system to cover a $300\text{ m} \times 300\text{ m}$ factory surface, the cost could be up to a half million US dollars. On the other hand, to deploy an active RFID system merely needs four RF readers and thousands of tags, which would cost less than 10,000 US dollars. Moreover, the deployment of RFID systems is more flexible than video camera systems due to the omnidirectional feature of RF signals.

1.2 Our RFID Configuration

After looking into the specifications of different available systems, we have chosen the Spider System manufactured by RF Code [5] to implement our activity tracking prototype.

The RF reader's operating frequency is 303 MHz. The reader also has an 802.11b interface to communicate with other machines. The detection range is set at 150 feet, and this range can be increased to 1000 feet with the addition of a special antenna. Each reader can detect tags within 2 s. Each RF tag is preprogrammed with a unique 7-character ID for identification by readers. Tags send their unique ID signals at random with an average of 2 s.

1.3 Our Contributions

The major contributions of this work are as follows.

First, we introduce a novel RFID application that uses an array of stationary RF tags to monitor activities in large fields. Differing from the traditional radio-based localization methods, our approach does not require the tracking objects to carry any transmitters or receivers, such as RF readers or tags.

Second, we model a data mining problem that is critical for the activity monitoring application using RFID. Although many attractive sequential pattern mining approaches have been proposed [6], [7], [8], [9], [10], [11], [12],

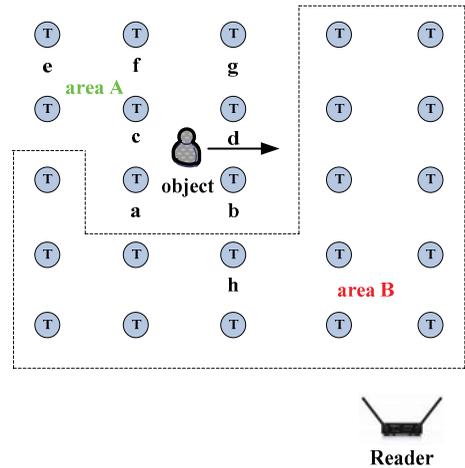


Fig. 2. Activity monitoring using RF tag arrays.

addressing the problem proposed in this paper is nontrivial, due to the noisy RF tag data. All the previous proposals assumed the data are precise, therefore, they cannot be applied to mining RF tag data. To solve the problem, we propose a fault-tolerant sequential pattern mining from an array of time series generated by the RF tags. Detail discussion on the challenge of this problem will be presented in Section 6.

Last, we conduct an empirical study using real RFID systems and data sets to verify the feasibility and the effectiveness of our approach. The experimental results show that the detection accuracy is perfect if we have appropriate parameters.

The rest of the paper is organized as follows: in Section 2, we describe our design of activity monitoring using RF tag arrays. We discuss the data collection and the preprocessing in Section 3 and present the frequent trajectory mining in Section 4. Our empirical study is reported in Section 5. Section 6 discusses the related work. We conclude the work in Section 7.

2 ACTIVITY MONITORING USING TAG ARRAYS

Most RFID applications attach RF tags to moving objects such as product items in a warehouse or customer carts in a store. In many scenarios, however, it is difficult to enforce an RF tag onto every object (e.g., people walking through the field).

To tackle this problem, instead of attaching one RF tag to each object, we propose to deploy an array of active RF tags onto the field. When an object moves through the field, the signals from some active tags will be affected and the RF readers will receive such signals. A database server collects the changes of signal strengths and uses the information to derive the activities in the field.

Fig. 2 illustrates this design, in which each hatched box is an RF tag. A set of RF tags are deployed on the field to be monitored. When an object (for example, a person in the figure) moves into the array, the signal strengths from some RF tags may change. In this example, the strengths from tags *a*, *b*, *c*, and *d* are very likely affected, while the signal strengths from the tags in area B, such as *h*, may not be affected.

Fig. 3 plots the signal strength changes of RF tags *c* and *h* on a real RF array deployment, as the one shown in Fig. 2.

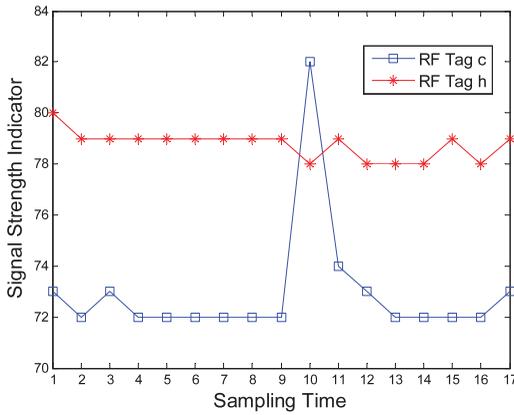


Fig. 3. Signal strengths of affected and unaffected RF tags.

The results indicate that when an object passes an RF tag such as c at time stamp 10, its signal strength is affected dramatically compared to an unaffected RF tag such as h .

By analyzing such changes, we want to derive the trajectories of the activities. Moreover, using the frequent trajectories, we can model the regular activities in a field. When an activity is detected, it can be compared with the frequent trajectories.

Due to the nature of RFID technology, we make the assumption that the number of simultaneous activities in a field is not large. For example, our method can detect several frequent trails that people walk along through a workshop. However, activities such as large parties in a hall or a banquet where hundreds of people walk about randomly cannot be handled well with our current method. Such situations can hardly be handled well by video monitoring systems either.

The novelty of our approach is that we use the interference on the RF tag signals caused by the activities to detect the activities of themselves or other unauthorized objects. However, it also poses the following two major challenges, which will be addressed in the remainder of this discussion.

Challenge 1: How to detect the positions of objects accurately. RFID data is very noisy. Tags often have very different characteristics [3]. Some RF tags are very sensitive, i.e., their signal is not stable even when no activities exist. The magnitude of the RF tags also varies. Different RF tags may give very different signal changes even if they are under the same interference.

Challenge 2: How to detect the frequent trajectories of activities. Since the RF tags are not synchronized in sending their signals, some activities may escape from one or a few tags. Moreover, since signals are not synchronized, the order of the changes may not correspond to the spatial-temporal order that an activity happens. How to detect the frequent trajectories effectively and efficiently is far from trivial.

3 DATA COLLECTION AND PREPROCESSING

Indeed, RF tags might respond differently to interference. In order to identify the interference from moving objects accurately, we need to capture the sensitivity of RF tags.

To measure the sensitivity, we first monitor the signal strengths of tags when no activity is present in the field for a

period of t . For each tag, we obtain a time series over the period. Let the set $\{s_1, s_2, \dots, s_t\}$ denote the signal strengths collected. We define the neutral value of the tag μ_s as the expected signal strength when there is no interference, i.e.,

$$\mu_s = \frac{\sum_{i=1}^t s_i}{t}.$$

The sensitivity of the RF tag is measured by the standard deviation of the time series, i.e.,

$$\sigma_s = \sqrt{\left(\sum (s_i - \mu_s)^2\right)/t}.$$

When an RF tag is used to detect activities and an object interferes with the signal of the tag, we call the activity an interference activity with respect to the tag. With the neutral value and the sensitivity of a tag, we can use a (small) number $k(k > 1)$ as the threshold to determine whether interference happens to a tag. Technically, we have the result below following from the Chebychev inequality.

Theorem 1 (Detection Threshold). Let μ and δ be the neutral value and the sensitivity of an RF tag, respectively. During the activity monitoring, if the reader receives a signal from the RF tag of strength s , and $|s - \mu| \geq k\sigma$, the probability that an inference activity happens is at least $(1 - \frac{1}{k^2})$.

Proof. Directly derived from Chebyshev's inequality. \square

We deploy an array of RF tags in a field. Each tag sends a signal in every unit period (called a period hereafter). RF tags are not synchronized. Instead, they compete for the transmission window. Thus, a tag may send its signal at the end of the period, and its neighbor tag may send its signal at the beginning of the period.

Several RF readers are connected to the server to collect signals. At the server side, a time series is accumulated for each tag and reader. Using the sensitivity and the neutral value of each tag, we transform the time series of a tag recorded by a reader R into a binary tag signal sequence (or tag sequences for short) s_i^R , where $s_i^R = 1$ if the tag is interfered in period i (i.e., $|s_i^R - \mu_s| \geq k\sigma$ according to Theorem 1), and $s_i^R = 0$ if the tag is not interfered in the period.

After the data collection and the preprocessing, we then use the tag signal sequences instead of the raw signal data in our data analysis.

4 FREQUENT TRAJECTORY MINING

In this section, we show how to mine frequent trajectories from the RF tag data. We first formulate the problem, and then introduce the algorithm.

4.1 Problem Formulation

Since the RF tags deployed are stationary, their spatial locations are known to the server. The data mining task consists of two phases: the training phase and the monitoring phase.

In the training phase, we collect the RF tag signal sequences over n periods, where n is a user specified length of time. In practice, the training period can be a day or a week, depending on the nature of the application. The

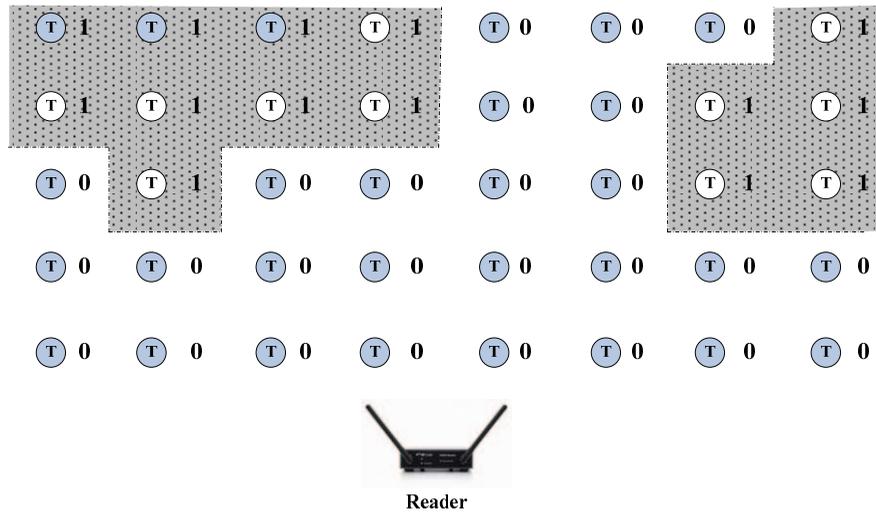


Fig. 4. Detecting borders.

sequences in the training phase will be used to find frequent trajectories as the model of the normal activities in the field.

In the monitoring phase, activities are detected and compared with the frequent trajectories. If an activity matches a trajectory, it is viewed as normal. Otherwise, an alert will be issued.

Since the trajectory matching is very similar to the approximate sequence matching problem, many existing methods can be used [1]. In the rest of the paper, we focus on the frequent trajectory mining problem (i.e., the training phase) only.

For each tag μ , let $s(\mu)$ be the tag signal sequence, and $s(\mu)_i$ be the signal in period i .

Intuitively, an activity can be described as a trajectory in the field under monitoring. In a period, the location segment of the object can be determined by the tags that are closest to the segment. Ideally, an activity can be captured by a series of RF tag sets $V_1 \rightarrow \dots \rightarrow V_l$, where $V_i (1 \leq i \leq l)$ is a set of RF tags describing the location segment of the object in period i , and the tag sets are interfered in consecutive periods.

If the tag sets can be detected accurately, the activity recognition problem is trivial. Due to the nature of RFID systems, however, there are a few important obstacles in practice.

First, not every RF tag along the trajectory may detect the activity. For example, in Fig. 2, if the object moves fast, it is possible that the object interferes with tag c but not tag d . Moreover, the probability that a tag fails to detect an activity is low but is unknown.

Second, the signals of tags may not accurately reflect the order of the activity. For example, in Fig. 2, although the object passes tag c before tag d , the interference may happen in the signal sequence of d before that of c . The reason is that the object may pass c right after c sends a signal of period i , but pass d right before d sends the signal in the same period. Therefore, the interference to d is reflected in period i , but the interference to c is recorded in period $(i + 1)$.

Third, an activity may interfere with multiple tags in a period. In order to derive the trajectories, we have to infer

the possible positions of the object based on the correlation of the interfered tags and the location of the readers.

In summary, the problem of mining frequent trajectory patterns from RF tag sequences is to explore the trajectories happening at least min_sup times in the training phase, where min_sup is a user specified frequency threshold.

4.2 Removing Redundancy and Detecting Borders

The RF tag signal collection has the following property.

Property 1. *If a reader R detects that an RF tag u is interfered in a period i , then for any RF tag v behind u in space with respect to R , with high probability, R detects v being interfered in at least one of the following periods: $(i - d)$, $(i - d + 1)$, i , $(i + d - 1)$, and $(i + d)$, where d is a user specified time shifting factor.*

Rationale. The property is clear in geometry, while it only holds with high probability, since if the object moves fast, there could be a slim window such that the signal of v is not affected. The probability is unknown and hard to be estimated. Thus, the property has to be used as a heuristic.

Using the above property, we can identify two types of redundancies among RF tag sequences. The first type is the redundancy among noninterfered tags. For example, in Fig. 2, all tags in area A are likely not interfered. We only need to know the area instead of individual values. The second type is the redundancy among interfered tags. For the same reason, the changes of tags e , f , and g in Fig. 2 are redundant.

To capture the activity in a period, the border between the interfered tags and the noninterfered tags is good enough. Thus, in each period and for each reader, we derive a border. The border detection works as follows: in a period i , we check $s(u)_i^R$ for each RF tag u and reader R . Recall that $s(u)_i^R$ is either 0 or 1. $s(u)_i^R$ is at the border if and only if there is at least one neighbor RF tag v such that $s(u)_i^R \neq s(v)_i^R$.

Fig. 4 illustrates the snapshot in a period for a reader. The borders are given by the dash-dot lines. The whitened boxes denote the borders of the interfered RF tags. There might exist cases that very few "0"s or "1"s appear inside of an "1" or "0" zone, so that these "0"s or "1"s are treated as outliers and will not be considered during the border detection.

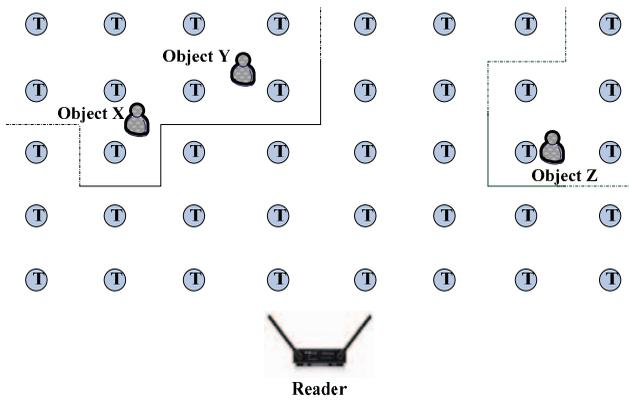


Fig. 5. The positions of objects.

Clearly, when the snapshot in period i can be held into main memory, the border detection takes $O(m)$ time, where m is the number of RF tags in the monitored field. Typically, m ranges from tens to thousands of RF tags, which can be easily accommodated in the main memory.

4.3 Identifying Possible Object Positions

Once we derive the borders between the interfered and noninterfered tags, we identify the possible locations of objects using the spatial map of the stationary tags.

Intuitively, the locations of objects are the outstanding parts of the border that a reader can see. For example, consider the case in Fig. 5. From the reader, two segments (the solid segments in the border) are the possible locations where objects exist. Heuristically, an object may appear proximate to an RF tag u if the tag is at the border and there is no other interfered RF tag blocking the connection between u and the reader, such as RF tags x , y , and z . By walking through the border once, we can identify the segments where an object may exist. We call such segments object location segments of the period w.r.t. the reader.

Please note that our location sensing is approximate. We only identify the ranges where objects may exist. Multiple objects may exist in the same range. In our trajectory mining algorithm, we shall use such ranges to assemble the possible trajectories. Another important issue is that some objects may hide behind other objects. For example, in Fig. 6, object B is hidden behind object A. Theoretically, we should be able to observe more degraded signals from the RF tags interfered by both A and B, such as the time shifting factor d . In the real system, however, the difference is often minor and not reliable for location detection.

To detect those hidden objects, we apply the following two methods.

First, we employ multiple readers. Multiple readers (e.g., 4-6) are deployed in a field so that the possibility that an object is hidden from all readers is reduced.

Second, we conduct fault-tolerant mining. As the objects are moving, one object hidden in one period may show up to some readers in other periods. As long as an object is not hidden at all times from all readers, our algorithm can detect the object.

4.4 The Mining Algorithm

The frequent trajectories are mined in the following two steps.

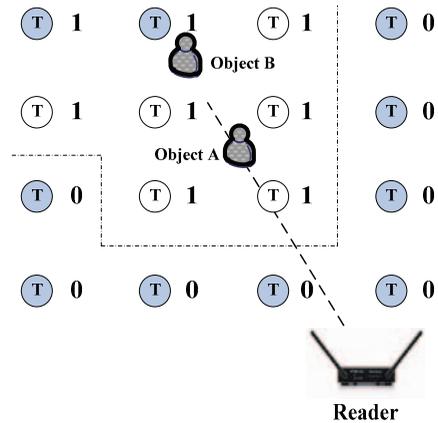


Fig. 6. Objects may be hidden.

4.4.1 Finding Frequent Positions of Objects

Clearly, a tag that is in an object location segment in a period is likely a part of the trajectory of an activity. The trajectory of a frequent activity may frequently trigger a tag in the object location segments. By scanning the object location segments in all periods once, we can find the tags that are in the segments in at least min_sup periods with respect to a reader.

Since an object can be occasionally hidden behind other objects, when counting the number of times a tag is in a segment, we also count the cases that tag is in the interfered side of the border. That is, if a tag is in the object location segments in some periods, and is interfered in some other periods, they are summed up together against the threshold min_sup .

We do not count the tags that are always hidden behind some tags in the object location segments. The rationale is that those tags are likely to be detected by other readers. On the other hand, if an activity is always hidden by some other activities, it is likely that either the activity is infrequent or it is a part of another activity. In many cases, the interfered tags not in the object location segments do not really capture the movement of objects.

<p>Input: RF tag signal sequences $\{s(u)_i^R\}$, frequency threshold min_sup</p> <p>Output: the set of frequent positions of objects w, r, t, reader R;</p> <p>Method:</p> <ol style="list-style-type: none"> 1: FOR each tag u DO <ul style="list-style-type: none"> create a counter $c_u = 0$ and a flag $f_u = 0$; 2: FOR each period i DO <ul style="list-style-type: none"> FOR each tag u DO 3: IF $s(u)_i^R = 1$ THEN $c_u = c_u + 1$; 4: IF u is at the border of interfered tags THEN $f_u = 1$; 5: FOR each tag u DO 6: IF $c_u \geq min_sup$ AND $f_u = 1$ THEN output u as a frequent position;

Fig. 7. Algorithm to find frequent positions of objects.

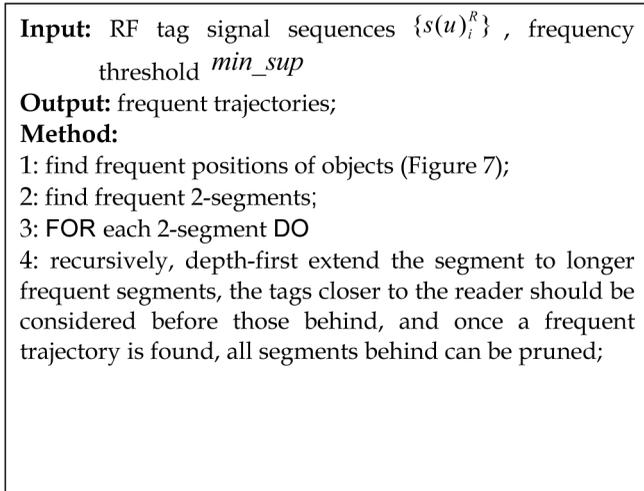


Fig. 8. The mining algorithm.

The method for finding the frequent positions of objects is illustrated in Fig. 7, in which we can see that the cost of the algorithm is one scan of the tag signal sequences. Thus, the complexity of our algorithm is $O(n)$, where n is the total number of tags.

4.4.2 Finding Frequent Trajectory Segments

As the second step, we find the frequent trajectory segments. The general idea is that we start with short segments and then use them to derive.

Conceptually, a l -segment of trajectory is a sequence $V_1 \rightarrow \dots \rightarrow V_l$ such that $V_j (1 \leq j \leq l)$ is a set of frequent positions of an object that are spatially adjacent, and V_q and $V_{q+1} (1 \leq q \leq l)$ are connected in space. In other words, the segment captures an activity in l periods such that V_j describes the trajectory of the activity in the j th period.

We start with finding 2-segments. We check the combinations of frequent object positions and examine whether they happen consecutively in space and in time. To tolerate faults, we allow some appearances in the reverse order. For example, if we see that tag a and tag b are interfered in consecutive periods frequently, and in some cases, b is interfered right before a , then, all those cases should be counted together as the support of $a \rightarrow b$. Technically, we use a threshold γ to specify the degree of fault tolerance. In a window of γ periods, the frequent positions can appear in any order. For example, if $\gamma = 2$, then $a \rightarrow b$ and $b \rightarrow a$ are considered matchable; if $\gamma = 3$, then $a \rightarrow b \rightarrow c$ and $c \rightarrow a \rightarrow b$ are matchable.

Typically, γ is a small positive integer such as 2 or 3. The proper value of γ depends on the maximal speed objects can move. If an object moves fast, it may have a better chance to cause more unsynchronized signals in more periods.

The space proximity is important here. It distinguishes the trajectories of consecutive movements from the spatial correlation of nonadjacent tags. Since a tag might be interfered by multiple moving objects, some tags nonadjacent in space may appear correlated. Those correlations should be filtered out in mining the frequent trajectories.

By scanning the tag signal sequences once, we can find all 2-segments and their counts (i.e., how many times a

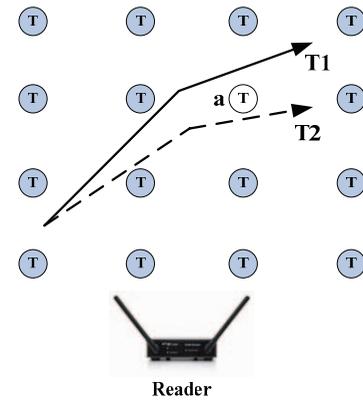


Fig. 9. Fault-tolerant mining.

segment appears in the training phase). Only those segments appearing at least min_sup times are retained as the frequent 2-segments, where min_sup is the frequency threshold.

Once the frequent 2-segments are found, we extend them to longer segments and check their support in the data set. To extend a frequent l -segment, we check all occurrences of the segment in the data set, and find the frequent positions in the next period following the segment. Those frequent positions adjacent in space form possible extensions to an $(l + 1)$ -segment. We check their frequency to identify the frequent $(l + 1)$ -segments. The extension of the frequent trajectory segments goes on until we cannot extend a frequent segment any more due to its frequency being lower than the threshold.

One important observation is that the same types of activities may not repeat their trajectories perfectly. For example, many people walk through a frequent trail, but each individual may have some variance. Fig. 9 shows such a case, where trails T_1 and T_2 should be considered as one type of activities following the same trajectory. T_1 does not interfere RF tag a while T_2 does. To handle such variance in the mining, we apply a fault-tolerant strategy based on Property 1 as follows.

We adopt a depth-first search to extend the frequent segments. The segments closer to the reader have a higher priority to be extended. Once a length $(l + 1)$ extension to V_{l+1} of a frequent l -segment $V_1 \rightarrow \dots \rightarrow V_l$ is infrequent, before we abort the extension, we check whether other extensions of the frequent segment are frequent. Particularly, we check those RF tags behind the tags in V_{l+1} . Fig. 8 summarizes the mining method.

5 EMPIRICAL STUDY

In this empirical study, we examine our frequent activities mining algorithm on a real implementation of 100 RF tags and 1 reader. As shown in Fig. 10 (only two rows are shown due to space limitations), these RF tags are deployed in 10 rows and each row has 10 RF tags in a field of size $10 \text{ m} \times 10 \text{ m}$. The distance between neighboring RF tags within a row or a column is 1 m. We let our student helpers to walk through this RF array following different routes and different speeds. The signal strength of each RF tag was recorded during the test period. By applying our mining

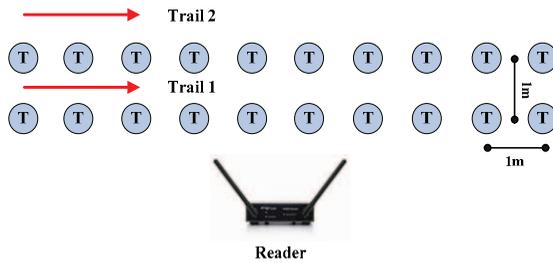


Fig. 10. Setup of Experiment 1.

algorithm on the readings of each RF tag received from the reader, we report the accuracy and efficiency of detecting trajectories of frequent activities.

To measure the detection accuracy, we use the ratio between the length of a correctly detected trajectory of frequent activities and the length of the real frequent route. We conduct six experiments, which represent common activities of people in large working areas, to estimate our algorithm. We start the tests with simple activities, such as single or consecutive activities with only one direction and one route for one object (Experiments 1 and 2), then we check the busy activities with multiple routes and directions (Experiments 3 and 4). Finally, we examine the complex activities with multiple objects and multiple trails (Experiments 5 and 6).

5.1 Experiment 1: Single Activity

The purpose of this experiment is to detect the trajectory of a single activity. We set up two routes (trails) in the RF array (as shown in Fig. 9). People walk through trails 1 and 2 independently for three times with different speeds (slow—0.5 m/s, fast—1.0 m/s).

The experimental results show that we can get 100 percent accuracy if we set the threshold, $min_sup = 2$, for detecting frequent positions of the objects, no matter what the walking speed of the people is. However, if we set the threshold $min_sup = 3$, the accuracy drops down to 60 percent. Due to the physical setting of RF tags, an RF tag sends a signal within a two second time frame, and there exist cases when people block an RF tag but this RF tag does not transmit any signals during the blocking period. Thus, the reader which fails to get the information about the RF tag was affected. As a consequence, this location may not be classified as a frequent one. Therefore, setting to a higher value may lead to

a lower accuracy. On the other hand, setting min_sup to a lower value may result in a large number of frequent locations and the computation cost of detecting frequent trajectories increases. We will test the effect of min_sup on detecting accuracy and efficiency in Experiment 3, where people may pass an RF tag many times during a busy activity.

5.2 Experiment 2: Group Activities

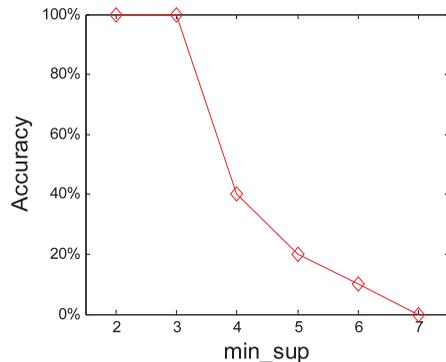
The purpose of this experiment is to find the trajectory of a temporally consecutive, group activity. We use the same setting as Experiment 1 and only select trail 1 for testing. We test the following scenario: one person walks through trail 1 at various speeds and the second person starts when the first one arrives at the 8th tag. All walks are in the same direction. In total, five people walk through the trail. We vary the people's walking speeds to test the robustness of the algorithm.

Again, the results indicate that our algorithm can detect the trajectory of a consecutive activity, trail 1, with 100 percent accuracy when we set the threshold of detecting frequent positions, $min_sup = 2$. We also test the case with $min_sup = 3$, and we find that we can still achieve 100 percent accuracy. This is because there are five consecutive objects passing the RF tags along the route. The results also show that the walking speed does not affect the detection accuracy as long as the activity is frequent.

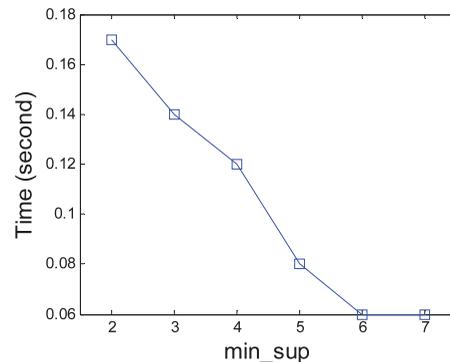
5.3 Experiment 3: Busy Activities

In this experiment, we test the capability of our method in detecting the trajectory of a busy activity. The same experiment setting of Experiment 1 is used here. We let one person walk back and forth on trail 1 at various speeds for one minute. Since the person may pass an RF tag many times during the one minute time period, we test the effect of min_sup (the threshold of frequent locations) on detection accuracy and efficiency, as shown in Fig. 11.

The results confirm what we discussed in Experiment 1. That is, with the increasing support threshold (min_sup in the figure), both the accuracy and the time cost are reduced. An interesting fact is that when $min_sup = 3$, we can achieve the best accuracy with the lowest time cost. Thus, how to set a proper value of support threshold for detecting frequent locations is an interesting work, which is left for our future investigation.



(a) Accuracy with respect to support threshold min_sup



(b) Runtime with respect to support threshold min_sup

Fig. 11. The effect of minimum support.

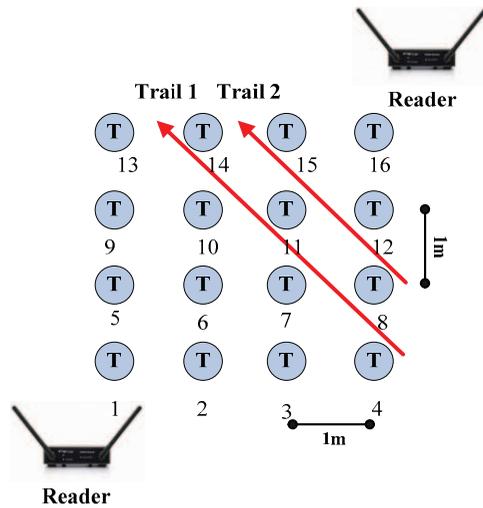


Fig. 16. Deployment of the Experiment 6.

5.6 Experiment 6: Multiple Trails

As previous discussion, the simple paralleled RFID array is hard to detect the real trail of a moving object. Therefore, we suggest an RFID grid deployed as Fig. 16 to enhance the accuracy of the trajectory detection. In this experiment, two people walk slowly following the different trails shown in Fig. 16.

Comparing all possible paths, we can obtain a boundary $8 \rightarrow 11 \rightarrow 14$. Other trajectories can be eliminated by using the outputs of two readers. However, another real trajectory ($15 \rightarrow 12$) was missed. In Fig. 17, the red solid line demonstrates the correct computed path which is one of real paths and the dashed line (blue line) denotes the possible trajectories.

For improving the performance of our algorithm, we attempt to deploy more readers in the sensing area. Some redundant patterns can be eliminated since we can obtain more information from extra readers. For example, one reader detects two patterns. One of them occurs at $time_a$ and another one appears at $time_a + t_1$. It is difficult to decide which pattern is the real trail, if those patterns are correlated to one position. Fortunately, at the same time, reader C also catches patterns related with this position. Based on the additional information, we can eliminate the illogical patterns.

5.7 Summary

Our empirical study using the RFID implementation confirms that using RF tags and readers to find trajectories of frequent activities is highly feasible. Our data mining techniques of mining fault tolerant frequent trajectories can detect frequent segments of activities. When the activities are not very complicated in space, the accuracy is high.

On the other hand, it remains a challenging task to improve the accuracy further for complex activities. We are working on using multiple readers for cross-validation as a promising solution.

6 RELATED WORK AND DISCUSSION

Sequential and approximate frequent pattern mining, and location sensing methods are highly related to this study.

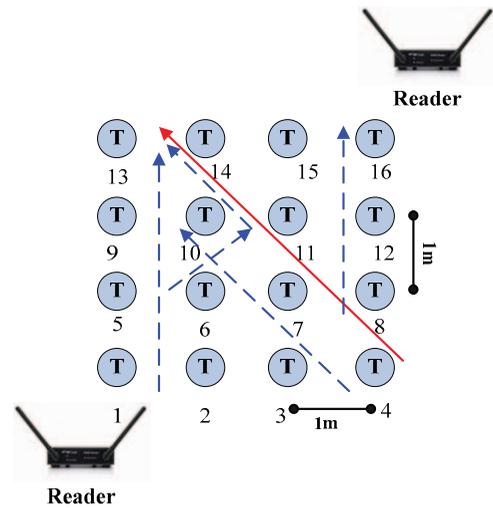


Fig. 17. Results of two readers.

6.1 Frequent Pattern Mining

Since it was first introduced [13], sequential pattern mining has been studied extensively. Conventional sequential pattern mining finds frequent subsequences in a sequence database based on exact match. There are two classes of algorithms. On one hand, the breadth-first search methods [2] are based on the a priori principle [14] and conduct level-by-level candidate-generation-and-tests. On the other hand, the depth-first search methods (e.g., PrefixSpan [15] and SPAM [16]) grow long patterns from short ones by constructing projected databases. Some variances of the depth-first search methods mine sequential patterns with vertical format [17]. Instead of recording sequences of items explicitly, they record item-lists, i.e., each item has a list of sequence-ids and positions where the item appears. As the real database may grow incrementally, researchers also propose incremental algorithms for the database to adaptively adopt new patterns [18].

Recently, Guralnik and Karypis used sequential patterns as features to cluster sequential data [19]. They project the sequences onto a feature vector comprised of the sequential patterns, and then use a k -means like clustering method on the vector to cluster the sequential data. Approximate frequent item set mining has also been studied [2]. Although the methods are quite different in techniques, they all explore approximate matching among item sets. For finding highly compact and discriminative patterns, Fan et al. propose a decision tree based approach to directly mine discriminative patterns as features vectors [6]. SwiftRule [20] utilizes the classification rules to conduct the time series mining to achieve easy-understood results for human experts.

From different point of view, Yang et al. presented a probabilistic model [17] to handle noise in mining strings. A compatibility matrix is introduced to represent the probabilistic connection from observed items to the underlying true items. Consequently, partial occurrence of an item is allowed and a new measure, match, is used to replace the commonly used support measure to represent the accumulated amount of occurrences. However, it cannot be easily generalized to apply on the sequential data targeted in this paper.

Chudova and Smyth used a Bayes error rate framework under a Markov assumption to analyze different factors that influence string pattern mining in computational biology [11]. Based on frequent sequence mining, Zaki et al. propose VOGUE [12], a variable order hidden Markov model, for modeling complex patterns in sequential data. Using the Time Series Knowledge Representation (TSKR) language, Moerchen proposes some mining algorithms for interval patterns expressing the temporal concepts of coincidence and partial order [21]. Recently, time series data is also used for the insight of system dynamics [22]. Extending the theoretical framework to mining sequences of sets could shed more light to the future research in this direction.

6.2 Location Sensing

Location sensing is a building block for many pervasive computing applications [23], [25], [26], [27]. Yossef et al. proposed the Device-free Passive localization (DfP) concept [28], which is similar to our basic idea [29]. They describe a prototype Wi-Fi systems and discuss potential challenges of DfP systems. TASA is a tag-free activity sensing framework, using passive tags [30]. Measurement model and the configuration of parameters are essential to DfP [31], [32]. By comparing the both the ideal case of signal dynamics and irregular information of moving objects, Zhang et al. [33] propose a real-time device-free tracking system with low latency. Different from the RSS-based DfP approaches, iLight uses light sensors and general light sources for localization [34]. Also, the device-free boundary coverage can be used for detecting intrusions [35].

On the other hand, Zhang et al. remark the link signature, such as RSSI and channel characteristics, for location distinction [36]. They present two approaches that are based on channel gains and channel impulse responses, respectively. The two approaches are combined with a complex temporal signature to discriminate location changes. The major problem of these approaches is that capturing the link signature is not trivial, especially for resource limited wireless devices, e.g., the RFID tag or sensors.

Trajectory pattern mining has been an important issue when deploying wireless sensors or RFID tags into physical space. Chen et al. focus on the problem of finding the k Best-Connected Trajectories (k -BCT) from a database such that the trajectories are geographically optimal for connecting the designated locations [37]. To predict complex movements, Jeung et al. propose a Hybrid Prediction Model, which estimates an object's future locations based on the recent movements and the pattern information [38]. The popularity of GPS provides effective trajectory representing solutions for people to quickly find their interesting places [39]. Lee et al. present a framework for frequent pattern-based classification [40]. Sequential patterns mining from time series is also employed in the Location-Based Service (LBS) [9]. Besides the localization of nodes, the boundary detection is also very important in the wireless networks, especially when location information is unavailable [41].

7 CONCLUSIONS

We propose to use RF tag arrays for activity monitoring. We present the framework, formulate the frequent trajectory

mining problem and develop a practical solution. Our empirical study using real RFID data sets verifies the effectiveness of the proposed method.

We are currently exploring the cross-validation method using multiple readers, and a more thorough test in real application fields. Moreover, it would be interesting to investigate the optimal deployment of RF tags and readers in a field. We will explore more applications of RFID technology in ubiquitous computing. Since RFID applications often generate a large amount of data, we believe those applications will pose new challenges and opportunities for data mining and pervasive computing research and development.

ACKNOWLEDGMENTS

Yunhao Liu's research is supported in part by the NSFC Distinguished Young Scholars Program under Grant 61125202, NSFC Major Program 61190110, and National High-Tech R&D Program of China (863) under grant No. 2011AA010100. Lei Chen is supported in part by RGC GRF under project No. 611608 and NSFC under project No. 60970112. Yiyang Zhao is supported by China Postdoctoral Science Foundation under project No. 2012M510454. Jian Pei's research is supported in part by an NSERC Discovery Grant, a BCFRST NRAS Endowment Research Team Program project, an SAP Business Objects ARC Fellowship, an NSERC CRD Research Grant, and a GRAND NCE project. Jinsong Han is supported by NSFC under project No. 61033015 and the Fundamental Research Funds for the Central Universities. All opinions, findings, conclusions, and recommendations in this paper are those of the authors and do not necessarily reflect the views of the funding agencies.

REFERENCES

- [1] X. Lian, L. Chen, J.X. Yu, G. Wang, and G. Yu, "Similarity Match over High Speed Time-Series Streams," *Proc. IEEE 23rd Int'l Conf. Data Eng. (ICDE)*, 2007.
- [2] J.K. Seppanen and H. Mannila, "Dense Itemsets," *Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining*, 2004.
- [3] L.M. Ni, Y. Liu, Y.C. Lau, and A.P. Patil, "LANDMARC: Indoor Location Sensing Using Active RFID," *Wireless Networks*, vol. 10, pp. 701-710, 2004.
- [4] K. Finkenzerler, *RFID Handbook: Fundamentals and Applications in Contactless Smart Cards and Identification*, second ed. Wiley, 2003.
- [5] RF Code, <http://www.rfcode.com/Products/Asset-Tags/Asset-Tags.html>, 2011.
- [6] W. Fan, K. Zhang, H. Cheng, J. Gao, X. Yan, J. Han, P. Yu, and O. Verscheure, "Direct Mining of Discriminative and Essential Frequent Patterns via Model-Based Search Tree," *Proc. 14th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining*, 2008.
- [7] J. Pei, J. Han, B. Mortazavi-Asl, H. Pinto, Q. Chen, U. Dayal, and M.-C. Hsu, "PrefixSpan: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth," *Proc. 17th Int'l Conf. Data Eng. (ICDE)*, 2001.
- [8] J. Yang, W. Wang, P.S. Yu, and J. Han, "Mining Long Sequential Patterns in a Noisy Environment," *Proc. ACM SIGMOD Int'l Conf. Management of Data*, 2002.
- [9] E.H.-C. Lu, V.S. Tseng, and P.S. Yu, "Mining Cluster-Based Temporal Mobile Sequential Patterns in Location-Based Service Environments," *IEEE Trans. Knowledge and Data Eng.*, vol. 23, no. 6, pp. 914-927, June 2011.
- [10] R. Agrawal and R. Srikant, "Mining Sequential Patterns," *Proc. 11th Int'l Conf. Data Eng. (ICDE)*, 1995.

- [11] D. Chudova and P. Smyth, "Pattern Discovery in Sequences under a Markov Assumption," *Proc. Eighth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining*, 2002.
- [12] M.J. Zaki, C.D. Carothers, and B.K. Szymanski, "VOGUE: A Variable Order Hidden Markov Model with Duration Based on Frequent Sequence Mining," *ACM Trans. Knowledge Discovery from Data*, vol. 4, no. 1, pp. 1-31, 2010.
- [13] R. Agrawal and R. Srikant, "Mining Sequential Patterns," *Proc. 11th Int'l Conf. Data Eng. (ICDE)*, 1995.
- [14] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules," *Proc. 20th Int'l Conf. Very Large Data Bases (VLDB)*, 1994.
- [15] J. Pei, J. Han, B. Mortazavi-Asl, H. Pinto, Q. Chen, U. Dayal, and M.-C. Hsu, "PrefixSpan: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth," *Proc. 17th Int'l Conf. Data Eng. (ICDE)*, 2001.
- [16] J. Ayres, J. Flannick, J. Gehrke, and T. Yiu, "Sequential Pattern Mining Using a Bitmap Representation," *Proc. Eighth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining*, 2002.
- [17] J. Yang, P.S. Yu, W. Wang, and J. Han, "Mining Long Sequential Patterns in a Noisy Environment," *Proc. ACM SIGMOD Int'l Conf. Management of Data*, 2002.
- [18] H. Cheng, X. Yan, and J. Han, "IncSpan: Incremental Mining of Sequential Patterns in Large Database," *Proc. 10th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining*, 2004.
- [19] V. Guralnik and G. Karypis, "A Scalable Algorithm for Clustering Sequential Data," *Proc. IEEE Int'l Conf. Data Mining (ICDM)*, 2001.
- [20] D. Fisch, T. Gruber, and B. Sick, "SwiftRule: Mining Comprehensible Classification Rules for Time Series Analysis," *IEEE Trans. Knowledge and Data Eng. (TKDE)*, vol. 23, no. 5, pp. 774-787, May 2011.
- [21] F. Moerchen, "Algorithms for Time Series Knowledge Mining," *Proc. 12th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining*, 2006.
- [22] P. Wang, H. Wang, and W. Wang, "Finding Semantics in Time Series," *Proc. Int'l Conf. Management of Data (SIGMOD)*, 2011.
- [23] Z. Zhong and T. He, "RSD: A Metric for Achieving Range-Free Localization beyond Connectivity," *IEEE Trans. Parallel and Distributed Systems*, vol. 22, no. 11, pp. 1943-1951, Nov. 2011.
- [24] M. Li and Y. Liu, "Rendered Path: Range-Free Localization in Anisotropic Sensor Networks with Holes," *IEEE/ACM Trans. Networking*, vol. 18, no. 1, pp. 320-332, Feb. 2010.
- [25] Y. Shang, W. Ruml, Y. Zhang, and M. Fromherz, "Localization from Connectivity in Sensor Networks," *IEEE Trans. Parallel and Distributed Systems*, vol. 15, no. 11, pp. 961-974, Nov. 2004.
- [26] Z. Yang, Y. Liu, and X. Li, "Beyond Trilateration: On the Localizability of Wireless Ad-Hoc Networks," *IEEE/ACM Trans. Networking*, vol. 18, no. 6, pp. 1806-1814, Dec. 2010.
- [27] Z. Yang and Y. Liu, "Quality of Trilateration: Confidence Based Iterative Localization," *IEEE Trans. Parallel and Distributed Systems*, vol. 21, no. 5, pp. 631-640, May 2010.
- [28] M. Youssef, M. Mah, and A. Agrawala, "Challenges: Device-Free Passive Localization for Wireless Environments," *Proc. ACM MobiCom*, 2007.
- [29] Y. Liu, L. Chen, J. Pei, Q. Chen, and Y. Zhao, "Mining Frequent Trajectory Patterns for Activity Monitoring Using Radio Frequency Tag Arrays," *Proc. IEEE Fifth Int'l Conf. Pervasive Computing and Comm. (PerCom)*, 2007.
- [30] D. Zhang, J. Zhou, M. Guo, J. Cao, and T. Li, "TASA: Tag-Free Activity Sensing Using RFID Tag Arrays," *IEEE Trans. Parallel and Distributed Systems*, vol. 22, no. 4, pp. 558-570, Apr. 2011.
- [31] X. Chen, A. Edelstein, Y. Li, M. Coates, M. Rabbat, and A. Men, "Sequential Monte Carlo for Simultaneous Passive Device-Free Tracking and Sensor Localization Using Received Signal Strength Measurements," *Proc. 10th Int'l Conf. Information Processing in Sensor Network (IPSN)*, 2011.
- [32] J. Wilson and N. Patwari, "A Fade Level Skew-Laplace Signal Strength Model for Device-Free Localization with Wireless Networks," *IEEE Trans. Mobile Computing*, vol. PP, no. 99, p. 1, 2011.
- [33] D. Zhang, Y. Liu, and L.M. NI, "RASS: A Real-Time, Accurate and Scalable System for Tracking Transceiver-Free Objects," *Proc. IEEE Int'l Conf. Pervasive Computing and Comm. (PerCom)*, 2011.
- [34] X. Mao, S. Tang, X. Xu, X.-Y. Li, and H. Ma, "iLight: Indoor Device-Free Passive Tracking Using Wireless Sensor Networks" *Proc. INFOCOM*, 2011.
- [35] A. Chen, S. Kumar, and T.H. Lai, "Local Barrier Coverage in Wireless Sensor Networks," *IEEE Trans. Mobile Computing*, vol. 9, no. 4, pp. 491-504, Apr. 2010.
- [36] J. Zhang, M.H. Firooz, N. Patwari, and S.K. Kasera, "Advancing Wireless Link Signatures for Location Distinction," *Proc. ACM MobiCom*, 2008.
- [37] Z. Chen, H.T. Shen, X. Zhou, Y. Zheng, and X. Xie, "Searching Trajectories by Locations: An Efficiency Study," *Proc. ACM Int'l Conf. Management of Data (SIGMOD)*, 2010.
- [38] H. Jeung, Q.L. Tasmanian, H.T. Shen, and X. Zhou, "A Hybrid Prediction Model for Moving Objects," *Proc. IEEE 24th Int'l Conf. Data Eng. (ICDE)*, 2008.
- [39] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma, "Mining Interesting Locations and Travel Sequences from GPS Trajectories," *Proc. World Wide Web (WWW)*, 2009.
- [40] J.-G. Lee, J. Han, X. Li, and H. Cheng, "Mining Discriminative Patterns for Classifying Trajectories on Road Networks," *IEEE Trans. Knowledge and Data Eng.*, vol. 23, no. 5, pp. 713-721, May 2011.
- [41] O. Saukh, R. Sauter, M. Gauger, P.J. Marrón, and K. Rothermel, "On Boundary Recognition without Location Information in Wireless Sensor Networks," *ACM Trans. Sensor Networks*, vol. 6, no. 3, pp. 1-35, 2010.



Yunhao Liu received the BS degree in automation from Tsinghua University, Beijing, China, in 1995, and the MS and PhD degrees in computer science and engineering from Michigan State University, in 2003 and 2004, respectively. He is the director of Key Laboratory for Information System Security, Ministry of Education. He is a professor at the School of Software, and a member of EMC Chair Professor Group at the Department of Computer Science, Tsinghua University. He is also a faculty member at the Department of Computer Science and Engineering, Hong Kong University of Science and Technology. His research interests include pervasive computing, peer-to-peer computing, and sensor networks. He is a senior member of the IEEE.



Yiyang Zhao received the BSc degree from Tsinghua University, China, the Mphil degree from the Institute of Electrical Engineering of CAS, China, and the PhD degree in the Department of Computer Science and Engineering from the Hong Kong University of Science and Technology (HKUST), China, in 1998, 2001, and 2010, respectively. His research interests include RFID, pervasive computing, distributed systems, embedded systems, and high-speed networking. He is a member of the IEEE and IEEE Computer Society.



Lei Chen received the BS degree in computer science and engineering from Tianjin University, China, the MA degree from Asian Institute of Technology, Thailand, and the PhD degree in computer science from the University of Waterloo, Canada, in 1994, 1997, and 2005, respectively. Now, he is working as an associate professor in the Department of Computer Science and Engineering at Hong Kong University of Science and Technology. His research interests include uncertain databases, graph databases, multimedia and time series databases, and sensor and peer-to-peer databases. He is a member of the IEEE.



Jian Pei received the PhD degree in computing science from Simon Fraser University in 2002. Currently, he is working as a professor at the School of Computing Science, Simon Fraser University. His research has been well recognized by several prestigious awards, such as several best paper and most influential paper awards from premier academic conferences, the 2005 British Columbia Innovation Council Young Innovator Award, and the IBM Faculty Award.

His research has been well funded by government funding agencies such as NSERC and US National Science Foundation (NSF). His strong connection with industry is reflected by his extensively funded projects by industry leaders such as Microsoft, IBM, HP, and SAP BusinessObjects. His research leads to not only numerous publications that have been cited more than ten thousands of times, but also to critical techniques that have been patented and adopted by the latest commercial products and in-house enterprise-wide data platforms. He has provided active services to international R&D professional communities. He is a senior member of the IEEE and the ACM.



Jinsong Han received the PhD degree in computer science and engineering from the Hong Kong University of Science and Technology in 2007. Currently, he is working as an associate professor at the Xi'an Jiaotong University. His research interests include peer-to-peer computing, anonymity, pervasive computing, network security, and high-speed networking. He is a member of the IEEE, IEEE Computer Society, and the ACM.

▷ **For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.**