# Anti-Localization Anonymous Routing for Delay Tolerant Network

*Xiaofeng Lu, +Pan Hui, ++Don Towsley, *Juahua Pu, *Zhang Xiong

*School of Computer Science
Beijing University of Aeronautics and Astronautics, Beijing, China
luxf@cse.buaa.edu.cn, pujh@buaa.edu.cn, xiongz@buaa.edu.cn
+Deutsche Telekom Laboratories and TU-Berlin,
Berlin, Germany
Pan.hui@telekom.de
+Dept. of Computer Science, University of Massachusetts
Amherst, U.S.A.
towsley@cs.umass.edu

December 16, 2009

**Abstract**

This paper focuses on the problem of how to allow a source to send a message without revealing its physical location and proposes an anti-localization routing protocol, ALAR, to achieve anonymous delivery in Delay/Disruption Tolerant Networks. The objectives of ALAR are to minimize the probability of a data source being localized and to maximize the destination's probability of receiving the message. ALAR can protect the sender's location privacy through message fragmentation and forwarding each segment to different receivers. ALAR is validated on two real-world human mobility datasets. This study indicates that ALAR increases the sender's anonymity performance by over 81% in different adversary densities with a 5% reduction in delivery ratio.

## 1 Introduction

In a Mobile Ad hoc Network (MANET), nodes communicate with each other from time to time and maintain dynamic and temporary connectivities through peer-to-peer wireless communication. If nodes move unpredictably at a high speed, disconnections between nodes can be frequent and a path between any node-pair may not be always possible. Such a kind of MANET is referred to as a Delay or Disruption Tolerant Network (DTN) [1].

Source-based routing techniques are inappropriate for DTN since the selected path (if any discovered) will most likely be invalid before it is used [1] [2]. Instead, nodes can transmit packets in a store-carry-forward fashion. They choose suitable encounter nodes as relays and forward their packets to these relays. When these relay nodes meet other nodes later, they forward the packets to the new relays. This packet delivery is analogous to the spread of infectious diseases [3] [4] [5]. This kind of routing is referred as *epidemic routing*. It can result in many replicas of the packet in the network, and once a copy of the packet reaches the destination node, the delivery is considered as successful [6] [7] [8]. An active area of research on routing in DTN focuses on minimizing the number of replicas while keeping delivery ratio high [9] [10].

The motivation of this paper is to provide location anonymity for nodes in an un-trusted wireless networks. This paper considers the adversarial localization as the major threat to the security of mobile wireless networks. Take the Iranian people for example. They do not have the right to criticize politicians on the Internet because of censorship, but still they can send their politic opinions by using wireless networks. The nature of shared transmission media makes wireless networks very vulnerable to security threats. The autocratic government may eavesdrop on the open-air wireless communication to detect sensitive messages and localize the senders. In an un-trusted network, authors do not want others to know their identities and positions for personal security reason. Thus, the physical location privacy is vital to the author of this message.

Here are the definitions of some terms being used in this paper.

- A sender is the author of an original message.
- A relay is a node that forwards packets to others.
- A transmitter is a node that transmits the electromagnetic wave. A transmitter can be a sender or a relay.
- A destination is the specific node the sender wishes to send a packet to.
- An ordinary node is a node that does not check the sensitivity of a message.
- An adversary is a node that checks the sensitivity of a message. Adversaries try to localize the sender of a sensitive message.

The contribution of this paper is the design of a novel Anti-Localization Anonymous Routing protocol for DTN called ALAR. The basic idea of ALAR is (1) to divide a message into $k$ segments and (2) to send each segment to $n$ different receivers. ALAR is designed to meet the following objectives:

- 1. Minimize the probability of a sender being localized by adversaries, $P_l$.
- 2. Maximize the number of receivers that gets the packet after time $t$, $I(t)$. As $I(t)$ depends on the network scale, we transform objective 2 into maximizing the ratio of receivers to all nodes after time $t$, which is called $P_r(t)$.

Some protocols have been proposed to provide the users location privacy at the application layer [16] [17]. The application cannot know user's location or trace user's movement under some conditions by these protocols. However, if the receivers localize the user using localization algorithm, they may get the user's location. Anti-localization anonymous routing in DTN is an open issue that has not been studied before.

This paper is organized as follows: Section 2 is the review of related works about anonymous routing protocols and secure routing protocols, and section 3 is the introduction of the localization algorithms. In section 4, the adversary model is introduced, and then in section 5, the anti-localization anonymous routing protocol, or ALAR, is introduced. In the following section, there is the validation of ALAR on human contact datasets and section 7 will be the conclusion of the whole paper.

# 2 Related Work

Many protocols have been proposed to provide anonymity for MANETs. This section reviews some related works about anonymous routing protocols and secure routing protocols for MANETs.

Hong et.al studied the relation between mobility and anonymity [11]. They presented an extensive study on new anonymity threats and classified the corresponding security requirements into three new categories: (1) venue anonymity, (2) privacy of ad hoc network topology, and (3) privacy of motion pattern. This paper focuses on ”venue anonymity”. Hong suggested to mix on-demand routing, identity-free routing and neighborhood traffic to generate new routing protocols for defending against threats in mobile networks. ALAR combines the ideas of on-demand routing, identity-free routing, neighborhood traffic, and physical localization to provide a better position privacy.

Zhang et.al proposed an anonymous on-demand routing protocol, MASK, for MANETs [12]. In MASK, nodes authenticate their neighboring nodes without revealing their identities to establish pairwise secret keys in a neighborhood authentication process. By utilizing the secret keys, MASK achieves routing and forwarding without disclosing the identities of participating nodes. However, adversaries in MASK can localize the sender's position.

ANODR is an anonymous protocol based on on-demand routing that provides route anonymity and location privacy to MANETs [14]. In terms of route anonymity, ANODR prevents adversaries from tracing a packet flow back to its source or destination; in terms of location privacy, ANODR ensures that adversaries cannot discover the real identities of local transmitters. However, the privacy ANODR provides is the identity of the sender instead of his/her physical location.

Some secure routing protocols, such as SEAD [15] and ARAN [23], employ authentication to ensure the receiver of packets is valid rather than compromised. These secure routing protocols try to protect the security of the contents of communication, but not the security of the sender. Authentication cannot fully thwart traffic analysis and localization algorithm.

Zhu et.al proposed a secure routing protocol ASR for MANETs [13]. Instead of encrypting the whole packet, they suggest encrypting a small piece of information and sending it together with the data packet. A relay node only needs to verify the small

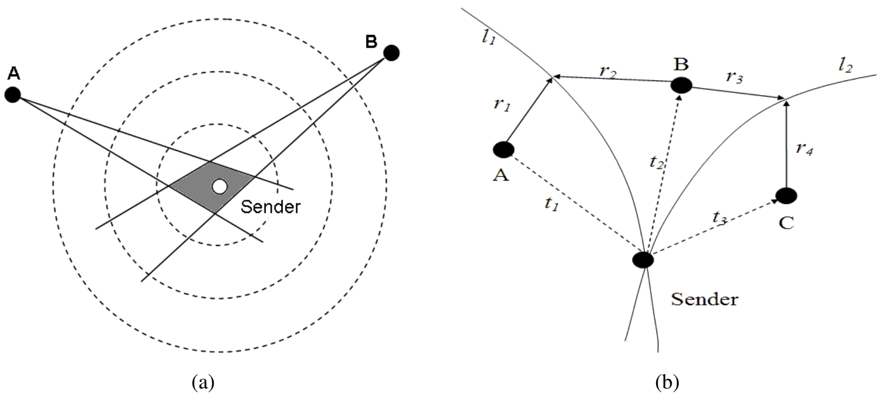


Figure 1: Illustrations of localization algorithms. (a) AOA localization algorithm and (b) TDOA localization algorithm, in which node A, B and C localize the sender

piece of information, rather than the whole packet. In ASR, Zhu's solution makes use of shared secrets between any two consecutive nodes. The goal of ASR is to hide the source and destination information from data packets rather than protect the source's physical location privacy.

In short, most secure routing protocols and anonymous routing protocols proposed for MANETs focus on ensuring that the receiver of packets is authenticated and the receivers or intruders cannot determine the identify of the sender. Few of these protocols prevent the adversary from localizing.

# 3 Preliminary: Localization Algorithm

If a node wants to acquire its location information when it does not have a GPS device, it can employ the localization algorithm to get its location from beacon nodes that know the location of themselves. Receivers can also execute the localization algorithm to acquire the sender's location information.

Most of the existing localization approaches fall into two categories: the range-based schemes and the angle-based schemes [18]. Range-based schemes rely on the range measurements, which can be achieved by computing the received signal strength (RSS), time of arrival (TOA) and time difference of arrival (TDOA). Angle-based schemes relay on the angle of arrival (AOA). The RSS is the easiest way to obtain, but the other types of measurements can provide much better accuracy. Localization algorithms using TDOA are discussed in [19], [20], [26], while angle measurements are exploited in [21], [22].

AOA algorithm requires directional antennas by which receivers can know the angle of arrival signals. With the directional antenna, at least two non-collinear neighboring receivers are required to discover the location of a sender as Figure 1 (a) shows. In Figure 1 (a), both A and B receive the sender's signals and know its direction, then A and B exchange their AOA measurements to calculate the sender's approximate location.

As AOA algorithm requires directional antennas to know the signal arrival angle, it is mostly available in small wireless networks. The most popular localization algorithms are range-based triangle localization algorithms. Triangle localization algorithm needs at least 3 beacon nodes to compute an unknown node's location. TDOA is a widely used range-based triangle localization algorithm. The main challenge of TDOA algorithm is clock synchronization. However, nodes can achieve clock synchronization using GPS. The precision of GPS data could be less than 10 nanoseconds. In fact, the principle of TDOA and other range-based triangle localization algorithms are the same.

TDOA algorithm is as the following:

1. Assume a sender sends a packet at time $t = 0$, and $h$ receivers receive it at different time $t_i(i = 1, 2, ..., h)$.

2. Receivers share their time-of-arrivals and compute differences in the time-of-arrivals (TD) of this packet, $TD = t_i - t_j(i \neq j)$.

3. Then receivers compute each corresponding spatial difference to the sender, $\triangle r_{i,j} = (t_i - t_j) \cdot C, (i, j = 1, 2, ..., h, i \neq j)$, where $C$ is the speed of light and $(i, j)$ is an enumeration of all pairs of receivers. Here, we assume each TD value is measured to a precision of about 10 nanoseconds which corresponds to about 3 meters.

4. With two receivers, they can get a curve on which any point has the same $\triangle r$ to the sender.

5. At least three receivers with known positions are required to find a 2D-position from two TDOAs as Figure 1 (b) shows.

# 4 Adversary Model

## 4.1 Passive Threat: Localization

Because of the nature of shared media in electromagnetic transmission, the electromagnetic wave is inevitable to be detected by detectors operated by eavesdroppers. Eavesdropping leads to passive type of attack. Active attacks would like to start route disruption or Denial of Service (DoS) attack. However, passive adversaries will try to be as invisible as possible, until it starts to destroy the sender physically. This kind of passive threat is hard to be detected, so it is another vital threat to MANETs [11].

It needs to be emphasized that a node's communication range depends on its transmission power, radio propagation, its antenna gain and the receiver's antenna gain, etc. If the distance from the eavesdropper to the sender is very long, the received transmission power at the eavesdropper's antenna is too weak to be distinguished from the noise because of the path loss and shadowing [24]. Therefore, we assume the eavesdropper's detection range is limited for the simplification reason.

With the help of localization algorithms, eavesdroppers can compute a transmitter's position. In addition to launching a network attack, the eavesdropper can launch a physical arrest if they know the transmitter's real location. We assume the adversary's detector is equipped with omni-directional antenna and executes TDOA algorithm to localize a transmitter. Suppose even the adversaries execute AOA algorithm, this study is still valid. We can change the definition of $P_l$ to there being at least 2 adversaries within the transmitter's transmission range.

### 4.2 Adversary Network

After an adversary receives a packet and certifies its content to be not sensitive, it will discard this packet. If the adversary odes not know a packet's content, it saves this packet into its buffer as a suspicious packet. The format of the suspicious record in the adversary's buffer is supposed to be (time stamp, packet id) where the time stamp is the system time when it receives this packet.

When the content of a message is large, the media access control protocol will cut the message into several small packets and send each of them respectively. In this condition, if an adversary only receives part of these packets, it will take these packets as suspicious packets because it is not able to know the content of the message. As an adversary is assumed to be mobile in the network, its radio device is supposed to be powered by batteries. If an adversary localized the transmitter whenever it receives a suspicious packet, the localization calculation would cost much of its limited batteries energy and CPU time. We assume an adversary localize a transmitter only after it knows the sensitivity of a packet.

When adversaries meet, they exchange their suspicious packets. Assume a message is cut into 2 packets and adversary A received packet 1 and adversary B received packet 2. After adversary A and adversary B exchanged their suspicious packets, both adversary A and adversary B have packet 1 and packet 2 and they know the content of this message. The adversaries do not help the sender and relays to forward packets but they exchange segments among adversaries.

## 5 Anti-Localization Anonymous Routing

### 5.1 Model and Assumptions

In this subsection, the models and assumptions which will be used throughout this paper are introduced.

- Transmission Model: Each node is equipped with a radio device by which a node can either transmit or receive packets, but not simultaneously. We assume these radio devices have the same transmission range. A node can only receive a packet when the transmitter of the packet is within its transmission range. Two nodes are called neighbors when they are within each other's transmission range. A radio device broadcasts a heartbeat signal periodically, which includes its identity. When a device receives a heartbeat signal from others, it knows that there is a neighbor and its identity. When the sender and relays wish to broadcast a packet, they set the destination address of this packet as the *Broadcast Address*.

  Before a node sends a segment, it collects heartbeat signals and takes the source nodes of these heartbeat signals as its neighbors. It sends a segment only when the following 2 conditions are met: (1) the number of heartbeat signals received

Table 1: An illustration of adversaries' buffer.

| adversary | A | adversary | B | adversary | C |
|---|---|---|---|---|---|
| time | packet id | time | packet id | time | packet id |
| $t_1$ | $packet_1$ | | | | |
| $t_2$ | $packet_2$ | $t_3$ | $packet_2$ | $t_4$ | $packet_2$ |

in the last period is larger than $n$, and (2) the receivers of previous segments do not overlap its current neighbors. Neither the sender nor the relay would send packets to a node twice.

- Energy Constrained: As is introduced in the precious section, an adversary's device is powered by batteries. If an adversary localizes all the transmitters of all suspicious packets, its batteries will be used up fruitlessly for there are lots of suspicious packets that are not sensitive. For this reason, we assume that an adversary computes the transmitter's position only after it has certified the content of this packet to be sensitive.

- Encryption and Decryption: It is reasonable to assume that the sender wishes the adversaries not be able to guess the content of a message before it receives all segments. Therefore, we assume that the sender might encrypt each segment and include the key of the decryption algorithm in the last segment. Hence, a receiver is able to know the content of a message only after it receives all segments of a message.

## 5.2 Protocol Description

As the sender does not know if there are adversaries within its neighborhood, it is dangerous to broadcast a sensitive message in one packet directly. The basic idea of ALAR is to split the original sensitive message into $k$ segments and send each segment to $n$ different neighbors. As relays transmit packets for several times, an adversary may receive several copies of a packet at different times from different transmitters. When they employ these suspicious records to localize the sender, they probably would not get the right answer.

For example, a sensitive message is divided into 2 packets. A relay $node_i$ forwarded $packet_2$ twice respectively at position 1 and position 2. Adversary A and B received $packet_2$ when $node_i$ sent it the first time at position 1, and adversary C received $packet_2$ when $node_i$ sent it the second time at position 2. The suspicious records of adversary A, B and C are as what Table 1 shows. When adversary A, B and C meet, they exchange their suspicious records and the three nodes know the content of this message. They run the TDOA localization algorithm to localize the transmitter. As adversary A, B and C received this packet from 2 transmissions at different places, they cannot get the right position of the transmitter.

The formal definition of ALAR is:

1. Assign two specified values to $k$ and $n$ according to network condition.

2. Divide a message into $k$ segments and encrypt each segment. The key of the decryption algorithm is in the last segment.

3. The sender sends each segment when (1) it has at least $n$ neighbors and (2) the receivers of $S_i$ do not overlap the receivers of $S_j$, $i \neq j$.

4. After a relay receives a segment from others, it forwards the segment to its neighbors when (1) it has at least $n$ neighbors and (2) the neighbors do not overlap the receivers of previous segments. A relay may forward a segment many times, and each time it forwards the segment that has been forwarded the least times.

Table 2 lists all the notation being used in the algorithm definition and the analysis later. The ALAR algorithm for the sender is:

```
The sender's receiver set is $X_o$, $X_o$ is empty
for $i$=1 to $k$ do
  while ($S_i$ not be sent) do
    Listen to the heartbeat signals
    if ($|Nb(sender)| \geq n$ and $Nb(sender) \notin X_o$) then
      The sender sends $S_i$
      $X_o \leftarrow Nb(sender)$
    end if
  end while
end for
```

The ALAR algorithm for a relay $node_j$ is:

```
$node_j$ has a receiver set $X_j$, $X_j$ is empty
loop
  Listen to the heartbeat signals
  if ($node_j$ receives a new segment $S_i$) then
    buffer← $S_i$
    Timer($S_i$)=0;
  end if
  if ($|Nb(node_j)| \geq n$ and $Nb(node_j) \notin X_j$) then
    Get a segment $S_k$ whose Timer($S_k$) is least
    $node_j$ sends $S_k$
    $X_j \leftarrow Nb(node_j)$
    Timer($S_k$)++
  end if
end loop
```

The key step of ALAR is to choose the appropriate value for $k$ and $n$ according to the network condition.

Table 2: Notation and Meaning.

| Notations | Meanings |
|---|---|
| $P_l$ | probability of being localized |
| $P_r(t)$ | the probability of the destination to receive the message after time t |
| $n$ | the threshold of neighbors |
| $k$ | the number of segments |
| $N$ | the number of ordinary nodes |
| $M$ | the number of adversaries |
| $L$ | length of experiment area |
| $S_j$ | segment $j$ |
| $\lambda$ | the density of adversaries, $\lambda = M/L^2$ |
| $v$ | the sender's speed |
| $v_1$ | the ordinary node's average speed |
| $V$ | relative speed of two nodes |
| $r$ | node's transmission range |
| $p$ | contact rate of nodes |
| $\|S\|$ | the number of nodes in set $S$ |
| $\\| * \\|$ | the number of adversary in the area $*$ |
| $Nb(node_i)$ | $node_i$'s neighbor set |

## 5.3 Probability Model

### 5.3.1 Preliminary: Epidemic Routing

Let $N$ be the total number of ordinary nodes moving within a square area $L^2$ and $M$ be the total number of adversaries. The density of adversaries is $\lambda$, $\lambda = \frac{M}{L^2}$. Now, we calculate $P_l$ and $P_r(t)$ respectively. First, we assume that a sender sends a sensitive message by epidemic routing [3] [4].

According to triangle location algorithm, $P_l$ is the probability of there being more than two adversaries within the sender's transmission range. Here, the probability of there being $i$ adversaries in the transmitter's communication range can be calculated by equation 1 according to spatial Poisson theory [27].

$$P(\|\pi r^2\| = i) = \frac{e^{-\lambda \pi r^2}(\lambda \pi r^2)^i}{i!},\ \lambda = \frac{M}{L^2} \tag{1}$$

where the notation $\|\pi r^2\|$ means the number of adversaries in the area $\pi r^2$. Then, $P_l$ is

$$\begin{aligned} P_l &= 1 - P(Not\ being\ localized) \\ &= 1 - P(\|\pi r^2\| < 3) = 1 - \sum_{i=0}^{2} \frac{e^{-\lambda \pi r^2}(\lambda \pi r^2)^i}{i!},\ \lambda = \frac{M}{L^2} \end{aligned} \tag{2}$$

According to [25], if a node's communication radius $r$ is largely smaller than the length of network area, say $r \ll L$, the rate $p$ at which a given node meets other nodes is

$$p = c\frac{Vr}{L^2} \tag{3}$$

where $c$ is a constant that depends on the mobility model used. We start from assuming the mobility model in our study to be the random direction model, $c = 1$. Assume the average velocity of ordinary nodes is $v_1$, the relative speed of two ordinary nodes is $V = \frac{c}{\pi v_1^2} \int_0^{2v_1} (\frac{x^2}{\sqrt{1-(\frac{2v_1^2-x^2}{2v_1^2})^2}})dx$ according to Groenevelt's research [25].

Equation 4 is the number of nodes that received a packet after time $t$ with one initial sender, where $I(t)$ represents the number of nodes received the packet, $p$ is the contact rate of the nodes and $t$ is the time duration from packets sending till present [4].

$$I(t) = \frac{N}{1 + e^{-pNt}(N-1)} \tag{4}$$

$P_r(t)$ is the ratio of receivers to all nodes after time $t$, so it can be got by equation 5.

$$P_r(t) = \frac{I(t)}{N} = \frac{1}{1 + e^{-pNt}(N-1)} \tag{5}$$

If we transform the objective 1 into maximizing the sender's probability of NOT being localized, which is refereed to as $1 - P_l$. Then the objective of our study is to maximize both $1 - P_l$ and $Pr(t)$. We define a new metric $CP(t)$ to determine the holistic performance.

$$CP(t) = (1 - P_l) \times P_r(t) = \frac{\sum_{i=0}^{2} \frac{e^{-\lambda\pi r^2}(\lambda\pi r^2)^i}{i!}}{1 + e^{-pNt}(N-1)} \tag{6}$$

### 5.3.2 ALAR

The transmitter by ALAR cannot send a segment as casually as by epidemic routing, so it would take longer time to send all segments compared to epidemic routing. Let the probability that a node has at least $n$ neighbors be $\varepsilon$.

$$\varepsilon = \frac{P(|\pi r^2| \geq n)}{\sum\limits_{i>0} P(|\pi r^2| = i)} \tag{7}$$

As $\sum\limits_{i\geq 0} P(|\pi r^2| = i) = 1$, $\sum\limits_{i>1} P(|\pi r^2| = i) = 1 - P(|\pi r^2| = 0)$. Then, equation 7 can be transformed into

$$\varepsilon = \frac{P(|\pi r^2| \geq n)}{1 - P(|\pi r^2| = 0)} = \frac{1 - \sum\limits_{i=0}^{n-1} \frac{e^{-x}x^i}{i!}}{1 - e^{-x}}, x = \frac{\pi r^2(N+M)}{L^2} \tag{8}$$

From the definition of ALAR we can know that the relays have fewer opportunities to forward a segment compared to epidemic routing. Let the ratio that receivers can receive a segment be $P_r^{(1)}(t)$ in ALAR and the ratio that they receive $k$ segments be $P_r^{(k)}(t)$.

$$P_r^{(1)}(t) = \frac{1}{1+e^{-p\varepsilon Nt}(N-1)} \tag{9}$$

$$P_r^{(k)}(t) = (P_r^{(1)}(t))^k = (\frac{1}{1+e^{-p\varepsilon Nt}(N-1)})^k \tag{10}$$

When the sender sends $S_k$, if there are more than two adversaries within its transmission range and at least one of these adversaries has received k-1 segments that the sender sent, these adversaries are able to know the content of this message immediately after they receive the last segment. If they certify the content of the message to be sensitive, they employ localization algorithm to localize the sender.

It is reasonable to assume that the sender cannot distinguish the adversary from the ordinary nodes, so the total nodes in the area that can receive messages is $(N+M)$. Adversaries can receive segments both from other adversaries and ordinary nodes. Assume the sender sends the segment $S_i$ at time $t_i$, the amount of nodes that have received the segment $S_i$ at time $t$ is

$$I_i(t) = \frac{(M+N)}{1+e^{-p\varepsilon(M+N)(t-t_i)}(M+N-1)} \tag{11}$$

As time increases, the number of nodes received segments increases accordingly. When the sender sends $S_k$ at time $t$, $I_{k-1}(t)$ is supposed to be smaller than $I_{k-2}(t)$,..., $I_1(t)$. So, we employ $I_{k-1}(t)$ as the approximate number of nodes that have received segments $S_1, S_2, ..., S_{k-1}$. Let the interval between $t_k$ and $t_{k-1}$ be $T$, $T = t_k - t_{k-1}$ and the area where $S_{k-1}$ can be carried to be A. Let the density of adversary nodes that have received segments $S_{k-1}$ be $\Lambda$ when the sender sends $S_k$ at time $t_k$.

$$\begin{aligned} \Lambda &= \frac{I_{k-1}(t_k)}{M+N}\frac{M}{A} = \frac{M}{A(1+e^{-p\varepsilon(M+N)T}(M+N-1))} \\ A &= Min(L^2, \pi(Tv_1+r)^2) \end{aligned} \tag{12}$$

The probability of the sender being localized by adversaries is the probability that there are at least three adversaries within the sender's transmission range and at least one of them has received all the previous segments, $S_1$, ...$S_{k-1}$.

$$\begin{aligned} P_l^{(k)} &= P(localized) = (1 - \sum_{i=0}^{2} \frac{e^{-\lambda\pi r^2}(\lambda\pi r^2)^i}{i!})(1-e^{-\Lambda\pi r^2}) \\ \Lambda &= \frac{M}{A(1+e^{-p\varepsilon(M+N)T}(M+N-1))},\ A = Min(L^2, \pi(Tv_1+r)^2) \end{aligned} \tag{13}$$

$$\begin{aligned} CP(t) &= (1-P_l^{(k)})P_r^{(k)}(t) \\ &= (\sum_{i=0}^{2} \frac{e^{-\lambda\pi r^2}(\lambda\pi r^2)^i}{i!})(1-e^{-\Lambda\pi r^2})(\frac{1}{1+e^{-p\varepsilon Nt}(N-1)})^k, \\ \Lambda &= \frac{M}{A(1+e^{-p\varepsilon(M+N)T}(M+N-1))},\ A = Min(L^2, \pi(Tv_1+r)^2) \end{aligned} \tag{14}$$

Table 3: The values of parameters.

| Parameters | Values | Parameters | Values |
|---|---|---|---|
| $N$ | 200 | $M$ | 1...100 |
| $r$ | 120 m | $L$ | 1000m |
| $v$ | 120m/minute | $v1$ | 60m/minute |
| $k$ | 2 | $n$ | 1...20 |

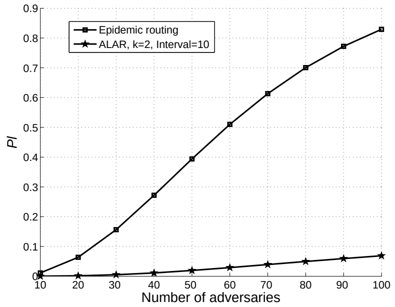


Figure 2: $P_l$ comparison with epidemic routing

Compare equation 13 with equation 2, we can know that $P_l^{(k)} < P_l$ because $(1 - e^{-\Lambda\pi r^2}) < 1$. This indicates that ALAR has better anonymity performance than epidemic routing.

## 5.4 Evaluation

In this section, we compare the $P_l$, $P_r(t)$ and $CP(t)$ of ALAR with that of the epidemic routing. The values of all parameters in this study are listed in table 3.

### 5.4.1 The Study of $P_l$

Figure 2 shows that with the increase of $M$, $P_l$ of epidemic routing increases sharply to 0.83 and $P_l$ of ALAR slowly increases to only 0.067. We can get a conclusion that fragmentation can evidently reduce the sender's probability of being localized by about $92\%$.

Figure 3 shows the variation of $P_l$ with $n$. In this case, the average number of neighbors is 9. When the value of $n$ is near to the average number of neighbors of a relay, $P_l$ increases rapidly. If $n$ is very small, the probability of there being more than 2 adversaries within the sender's transmission range is low. So the probability $P_l$ is low correspondingly. Besides, a relay has few chances to have lots of neighbors at one moment, so it has few opportunities to send segments when $n$ is large and the probability $P_l$ is very low inevitably.

Figure 4 shows the impact of the interval time $T$ on $P_l$. The adversaries that receive $S_{k-1}$ from the sender are not able to move far away in a short time, so they have a high

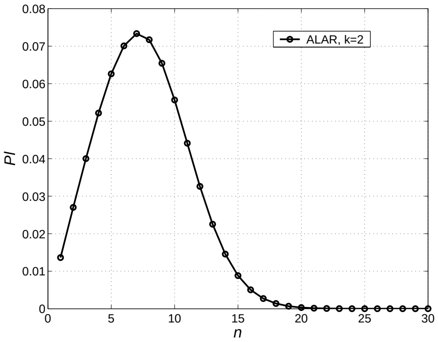


Figure 3: The impact of $n$ on $P_l$

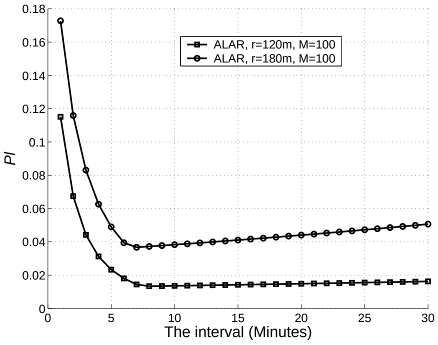
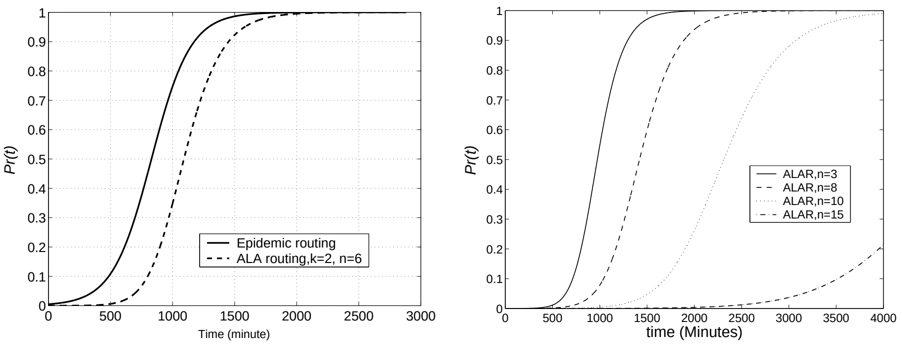


Figure 4: The impact of $interval$ on $P_l$

probability to receive the next segment $S_k$ from the sender if the interval between $t_k$ and $t_{k-1}$ is very short. Thus, if the interval between $t_k$ and $t_{k-1}$ is very short, $P_l$ is inevitably large. With the interval $T$ increases from 1 to 8 minutes, $P_l$ decreases by over 97%. $P_l$ is lowest when the interval is from 7 to 10. When the interval is larger than a threshold, with the increase of the interval, more nodes can receive $S_{k-1}$ in the interval and $\Lambda$ increases as well. Therefore, $P_l$ will increase with the increase of the interval. This figure also indicates that when other conditions are the same, the larger the sender's transmission range is, the larger $P_l$ is. Suppose nodes' transmission range $r$ is extremely large, the receivers can receive the sender's segments wherever it moves so as to the sender will be localized with the probability 100%.

### 5.4.2 The Study of $P_r(t)$

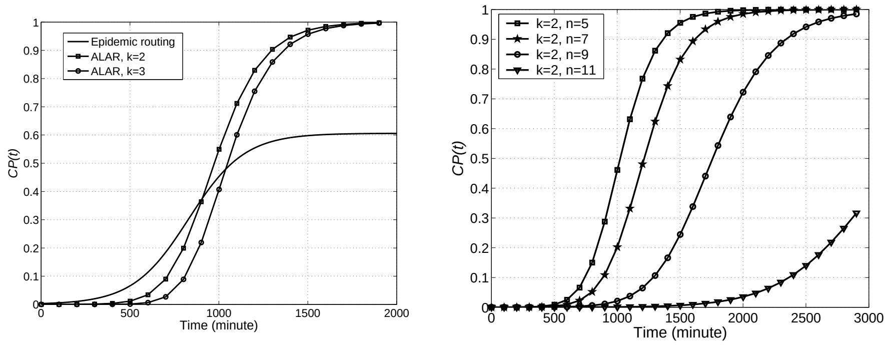
Figure 5 (a) shows the influence of the fragmentation on $P_r(t)$. The x-axis of it is the experiment time from when the sender begins to send segments, and y-axis of it is the ratio of receivers to all nodes after time $t$. ALAR defers the spreading of a message and it would take nodes longer time to receive all segments. It also indicates that ALAR would not decrease $P_r(t)$ in a longer term. Figure 5 (b) shows that the larger the $n$



(a)

(b)

Figure 5: (a) $P_r$ comparison with epidemic routing and (b) $P_r$ vs. $n$



(a)

(b)

Figure 6: (a) $CP$ comparison with epidemic routing and (b) $CP$ vs. $n$

is, the slower the network nodes receive all segments. The reason of this is that the smaller the $n$ is, the more opportunities a relay can forward segments. A node has few chances to have lots of neighbors at one moment. Therefore, it would take the relays much longer time to wait for a chance to forward a segment to others when $n = 10$ rather than when $n = 5$.

### 5.4.3 The Study of $CP(t)$

As maximizing $CP(t)$ is our final objective, we compare $CP(t)$ of epidemic routing and that of ALAR. Figure 6 (a) shows $CP(t)$ of ALAR is smaller than that of epidemic routing in the beginning phase because $P_r(t)$ of ALAR is much smaller than that of epidemic routing in that phase as Figure 5 (a) indicates. However, after certain time, $CP(t)$ of ALAR increases quickly and becomes larger than that of epidemic routing. This indicates that ALAR has better anonymity performance in sending messages in the system that can tolerate certain delay. It also shows the impact of $k$ on $CP(t)$. It takes other nodes longer time to receive all 3 segments than to receive 2 segments.

Figure 6 (b) illustrates the influence of $n$ on $CP(t)$. With the increase of $n$, it takes other nodes more time to receive all segments. The variation of $CP(t)$ is the same as the variation of $P_r(t)$ in Figure 5 (b). As $CP(t) = (1 - P_l) \times P_r(t)$ and $P_l$ is fixed, the variation of $CP(t)$ follows the variation of $P_r(t)$.

### 5.5 Discussion

From the study we have done, we know that ALAR can lower the sender's probability of being localized by adversaries. However, cutting a message into segments would inevitably correspond to longer delivery delay. Actually, most of the routing protocols for delay tolerant networks are not appropriate for instant communication. ALAR may induce longer delivery delay, but it can be employed to send or broadcast a message without the response from the destination and time constraint. A similar application is to publish an article on the Internet. When a user publishes an article on a Website, he/she does not know who will read this article and when they will read it. For the user, it does not matter even others read it after some days. Here, if we set $n = 1$, relays have more opportunities to forward a segment and the delivery delay would reduce, but this would cost relays much more battery energies.

A key step of ALAR is to choose the appropriate values of $k$ and $n$. We can know from equation 10 that with the increase of $k$, it takes longer time for $P_r(t)$ to achieve 1. So we suggest $k = 2$ for less delivery delay. From the study of the impact of $n$ on $P_l$, we know that $P_l$ decreases with the increase of $n$ when $n$ is larger than a threshold, but the delivery delay increases with the increase of $n$. Therefore, we suggest the appropriate value of $n$ to be from 2 to 4.

## 6 Validation

As human mobility plays a key role in packet delivery in DTN [29], we need to check the situation of user mobility in the real world. We chose to evaluate our protocol on two real-world experiment datasets to determine the impacts of human mobility on the routing of packets.

### 6.1 Mobility Analysis

In this study, we use the experimental dataset gathered at the IEEE *Infocom* 2005 conference by the Haggle Project (www.haggleproject.org) [28]. In the experiment, the device used to collect connection data was the Intel iMotes that had the same transmission and reception range. Each participant carried a iMote that logged the beginning time and the end time of any contact with other nodes and the device's id. The format of this dataset is *(i, j, $t_b$, $t_e$)*, where $t_b$ is the beginning time of a contact and $t_e$ is the end time of this contact. We define a variable *contact duration* to study the relative mobility between nodes, $contact\ duration = t_e - t_b$.

Figure 7(a) shows the distribution of the number of neighbors. We conclude that each node has at least one neighbor node with around 30 percent experiment time

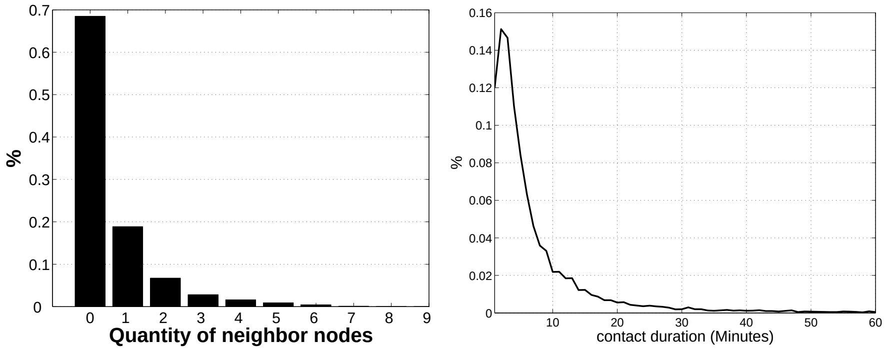


Figure 7: (A) Distribution of the quantity of neighbor nodes. (B) Distribution contact durations.

during the 4 days experiment time. This figure also shows that a node usually does not have too many neighbors. It is almost zero probability that a node has more than 7 neighboring nodes at one moment in this experiment.

Figure 7 (b) shows the distribution of contact durations between nodes. The statistical study of these contacts shows that more than 80 percent of the contacts are shorter than 10 minutes and more than 90 percent contacts are shorter than 20 minutes. This demonstrates that two nodes did not remain in contact for a long time. This is the feature of DTN.

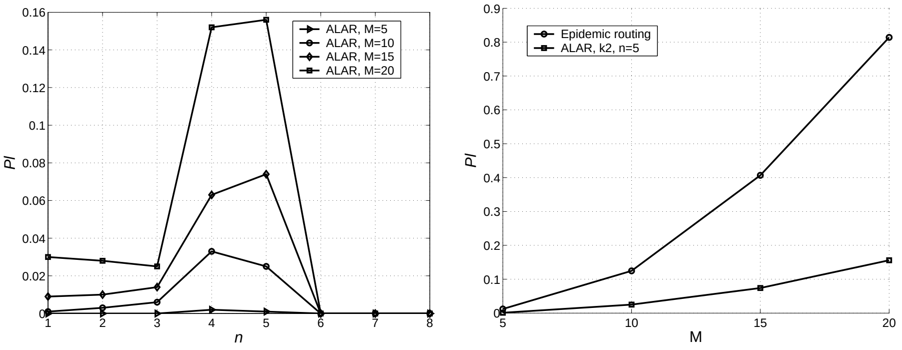
## 6.2 Simulation Results

### 6.2.1 Simulation Setup

According to the discussion about $n$, $n$ is 4 in this case. We perform a packet routing simulation with different $n$ on the *Infocom* dataset to study the impact of $n$. We compare the anonymity performance, delivery performance of ALAR with that of epidemic routing. To get the average $P_l$ and $P_r(t)$, we run the simulation program 1000 times with each $n$. In each simulation, we randomly select two nodes as the sender and the destination and randomly select 50% nodes as the adversaries.

### 6.2.2 Anonymity Performance: $P_l$

Figure 8 (a) shows the impact of $n$ on the sender's $P_l$. When $n$ is lower than 4, $P_l$ of ALAR is lower than 0.03 even when the proportion of adversary among all nodes is about $40\%$. However, $P_l$ increases sharply when $n$ is 4 or 5. The reason is that if the sender sends a segment only when it has 1 to 3 neighbors, the probability of there being at least 3 adversaries in its communication range is inevitably small. If the sender sends a segment when it has more than 3 neighbors, the probability of there being more than 2 adversaries within its communication range will increase. However, when $n$ is larger than 5, $P_l$ decreases to 0. As Figure 7 (a) indicates, a node has few chances



(a)

(b)

Figure 8: $P_l$ on the Infocom datasets. (a) $P_l$ vs. $n$ under different $M$ (b) $P_l$ comparison with epidemic routing

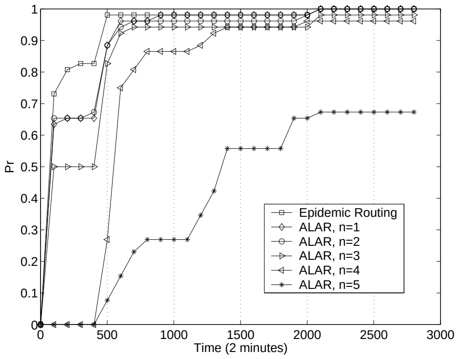


Figure 9: $P_r$ on the Infocom datasets

to have more than 6 neighboring nodes at one moment, so the sender has almost zero opportunity to send segments when $n \geq 6$.

In 8 (a), $P_l$ of ALAR is largest when $n = 5$, which means the worst anonymity performance of ALAR. Therefore, we compare the worst anonymity performance of ALAR with the normal anonymity performance of epidemic routing. Figure 8 (b) shows that with the increase of adversaries, both $P_l$s of ALAR and epidemic routing increase correspondingly. The increasing rate of $P_l$ of epidemic routing is much faster than that of ALAR. If there are 20 adversaries in the network, the sender's probability to be localized is larger than 0.8 if it employs epidemic routing, but $P_l$ of ALAR is just about 0.15. This statistical study also validates that ALAR can increase the sender's anonymity performance by at least 81% with 50% networks nodes being adversaries.

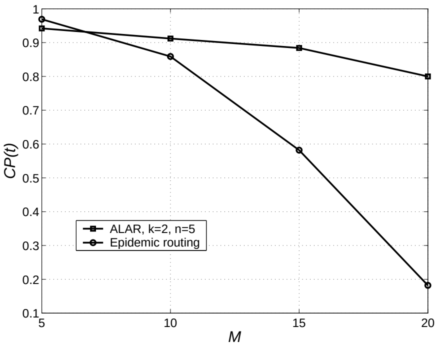


Figure 10: $CP(t)$ on the Infocom datasets

### 6.2.3 Delivery Performance: $P_r(t)$

Figure 9 shows $P_r(t)$ of ALAR and epidemic routing. The increases of $P_r(t)$ of ALAR when $n < 3$ is close to the increase of epidemic routing. When $n < 5$, $P_r(t)$ of ALAR after one day is almost the same as that of epidemic routing. With the increase of $n$, the delivery delay increases as well. It validates the correctness of the math model for the $P_r(t)$.

There are four durations during which $P_r(t)$ remains unchangeable. The four durations can be clearly found from the curve of $n = 5$. The four durations were four nights of Infocom'05 conference. As most of the participants did not meet others at nights, the forwarding process almost stopped and $P_r(t)$ remained stable. Still there were certain number of participants who met each other at night, so the delivery process could slowly carry on by epidemic routing.

### 6.2.4 Holistic Performance: $CP(t)$

Figure 10 shows $CP(t)$ of ALAR and epidemic routing after 1440 minutes with different numbers of adversary. It shows that when the number of adversary is 5, $CP(t)$ of epidemic routing is slightly higher than that of ALAR. When the number of adversary is low, a node has few opportunities to have more than 3 adversaries within its transmission range. Therefore the sender has very small probability to be localized even when using epidemic routing. Thus, the advantage of ALAR is not clear. However, with the increase of adversaries, $CP(t)$ of epidemic routing decreases rapidly and becomes much lower than that of ALAR soon. When the number of adversaries is 20, $CP(t)$ of epidemic routing after 1440 minutes is only about 22% of that of ALAR.

## 6.3 Experiment 2

We performed another validation simulation on the mobility dataset collected in *Cambridge, UK*. In this experiment, the iMotes were distributed mainly to two groups of students from University of Cambridge Computer Laboratory, specifically undergraduate year 1 and year 2 students [28]. This dataset lasted for 11 days. We checked the

Table 4: $P_l$, $P_r(t)$ and $CP(t)$

| $k$ | $n$ | $P_l$ | $P_r$ | $CP$ |
|---|---|---|---|---|
| 2 | 2 | 0.027 | 0.45 | 0.438 |
| 2 | 4 | 0.038 | 0.64 | 0.616 |
| 2 | 7 | 0.048 | 0.37 | 0.352 |

experiment dataset over 30 minutes and found the average number of neighbors is 6.2. We assign $n = 4$ and $k = 2$ in this case. We also checked $P_l$, $P_r(t)$ and $CP(t)$ with $n = 2$ and $n = 7$. The simulation results are listed in table 3. It shows that $CP(t)$ achieved the largest value when $n = 4$. It also verifies that the conclusion about the value of $n$ in section 5.6 is reasonable.

## 7 Conclusion

Location privacy is an important issue yet has not been well touched in DTN. In this paper, we start looking at this issue by introducing an anti-localization routing protocol, ALAR, which uses a series of *divide*, *forward*, and *move* procedures to increase node location privacy. We set up a probability model to compute the probability of the sender being localized by adversaries. This probability model shows ALAR can lower the sender's probability of being localized. We did validation simulations on two real-world mobility traces to certify the advantage of ALAR. These validation simulations show that ALAR can decrease the sender's probability of being localized by at least 81% with about 5% lose in $P_r$. Therefore, ALAR can increase the holistic anonymity performance by 77%.

We do not claim that we have found an optimum solution for location anonymous communication in DTN. In our current study, we did not consider the energy consumption and a relay can forward a segment many times. In the future, we will consider how to assure the delivery performance with limited delivery times and energy consumption.

## Acknowledgment

We are extremely grateful to the ITA Projects This work is also partly supported by National Natural Science Foundation of China (60803120), the PhD Innovation Foundation of Beihang University and Deutsche Telekom Laboratories Lens Projects.

## References

[1] K. Fall. Delay tolerant networking architecture for challenged internets, in Proceedings of SIGCOMM '03: 2003 conference on Applications, technologies, architectures, and protocols for computer communications, 2003.

[2] S. Burleigh, A. Hooke, L. Torgerson, K. Fall, V. Cerf, B. Durst, K. Scott, and H. Weiss. Delay-tolerant networking: An approach to interplanetary internet, in Communications Magazine, June 2004, vol.41, pages 128-136.

[3] X. Zhang, G. Neglia, J. Kurose, and D. Towsley. Performance modeling of epidemic routing, in Computer Networks, 2006, pages 2867-2891

[4] T. Small and Z. J. Haas. The shared wireless infostation model: a new ad hoc networking paradigm (or where there is a whale, there is a way), in Proceedings of MobiHoc'03: the 4th ACM international symposium on Mobile ad hoc networking & computing, 2003, pages 233-244.

[5] A. Vahdat and D. Becker. Epidemic routing for partially connected ad-hoc networks, Tech. Rep. Duke CS-2000-06, Duke University, April 2000.

[6] A. Panagakis, A. Vaios, and I. Stavrakakis. On the effects of cooperation in dtns, in Proceedings of COMSWARE: the Second IEEE/Create-Net/ICST International Conference on COMmunication System softWAre and MiddlewaRE, 2007.

[7] T. Spyropoulos, K. Psounis, and C. Raghavendra. Efficient Routing in Intermittently Connected Mobile Networks: The Multi-copy Case, in IEEE/ACM Transactions on Networking, 2008, pages 77-90.

[8] W. Zhao, M. Ammar, and E. Zegura. A message ferrying approach for data delivery in sparse mobile ad hoc networks, in Proceedings of ACM MobiHoc'04: the 5th ACM international symposium on Mobile ad hoc networking and computing, 2004.

[9] K. Psounis, T. Spyropoulos and C. Raghavendra. Single-copy routing in intermittently connected networks, in Proceedings of IEEE SECON: Society Conference on Sensor and Ad Hoc Communications and Networks, 2004.

[10] W. Zhao, M. Ammar, and E. Zegura. Controlling the mobility of multiple data transport ferries in a delay tolerant network, in Proceedings of INFOCOM'05: the 24th Annual Joint Conference of the IEEE Computer and Communications Societies, 2005.

[11] Xiaoyan Hong, Jiejun Kong, Mario Gerla. Mobility changes anonymity: new passive threats in mobile ad hoc networks: Research Articles, in Wireless Communications & Mobile Computing, 2006, pages 281-293

[12] Yanchao Zhang, Wei Liu, and Wenjing Lou. Anonymous Communications in Mobile Ad Hoc Networks, in Proceedings of IEEE INFOCOM'05, 2005.

[13] Bo Zhu, Zhiguo Wan, Mohan S. Kankanhalli, Feng Bao, and Robert H.Deng. Anonymous secure routing in mobile ad-hoc networks, in Proceedings of LCNP'04: the 29th Annual IEEE International Conference on Local Computer Network, 2004, pages 102-108.

[14] J. Kong and X. Hong. ANODR: ANonymous On Demand Routing with Untraceable Routes for Mobile Ad-hoc Networks, in Proceedings of ACM MOBIHOC'03, 2003, pages 291-302.

[15] YC. Hu, D. B. Johnson, and A. Perrig. SEAD: Secure Efficient Distance Vector Routing in Mobile Wireless Ad Hoc Networks, in Proceedings of WMCSA'02: the 4th IEEE Workshop on Mobile Computing Systems and Applications, 2002.

[16] Beresford, A.R. and F. Stajano. Location Privacy in Pervasive Computing, in IEEE Pervasive Computing Magazine, 2003, pages 46-55.

[17] Kido, H., Y. Yanagisawa, and T. Satoh. An Anonymous Communication Technique Using Dummies For Location-based Services, in Proceedings of IEEE ICPS'05: International Conference on Pervasive Services, 2005, pages 88-97.

[18] R. Peng and M. Sichitiu. Angle of arrival localization for wireless sensor networks, in Proceedings of IEEE SECON: 3rd Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks, 2006.

[19] N. Priyantha, A. Chakraborthy, and H. BALARkrishnan. The cricket location-support system, in Proceedings of International Conference on Mobile Computing and Networking, 2000, page 32-43.

[20] A. Savvides, C. C. Han, and M. B. Srivastava. Dynamic fine-grained localization in ad-hoc networks of sensors, in Proceedings of Mobicom'01, 2001, pages 166-179.

[21] D. Niculescu and B. Nath. Ad hoc positioning system (APS) using AOA, in Proceedings of IEEE INFOCOM'03, 2003.

[22] A. Nasipuri and K. Li. A directionality based location discovery scheme for wireless sensor networks, in First ACM International Workshop on Wireless Sensor Networks and Applications, 2002.

[23] K. Sanzgiri, B. Dahill, B. N. Levine, C. Shields and E. Royer. A Secure Routing Protocol for Ad Hoc Networks, in Proceedings of IEEE ICNP'02: the 10th International Conference on Network Protocols, 2002.

[24] Xiaofeng Lu, Fletcher Wicker, Don Towsley, Zhang Xiong, Pietro Lio. Detection Probability Estimation of Directional Antennas and Omni-directional Antennas, Wireless Personal Communication, 2009

[25] R. Groenevelt. Stochastic models in mobile ad hoc networks, Ph.D. dissertation, University of Nice Sophia Antipolis, April 2005.

[26] Fredrik Gustafsson. Positioning using time-difference of arrival measurements, in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2006, pages 553-556.

[27] Athanasios Kottas and Bruno Sans, Bayesian Mixture. Modeling for Spatial Poisson Process Intensities, with Applications to Extreme Value Analysis, 2006

[28] Pan Hui and Jon Crowcroft and Eiko Yoneki. BUBBLE Rap: Social-based Forwarding in Delay Tolerant Networks, in Proceedings of ACM MobiHoc'08: the 9th ACM international symposium on Mobile ad hoc networking & computing, 2008

[29] 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 Proceedings of IEEE INFOCOM'06, 2006.