Anti-localization anonymous routing for Delay Tolerant Network

Xiaofeng Lu\textsuperscript{a,}\*, Pan Hui\textsuperscript{b}, Don Towsley\textsuperscript{c}, Juahua Pu\textsuperscript{a}, Zhang Xiong\textsuperscript{a}

\textsuperscript{a}School of Computer Science, Beijing University of Aeronautics and Astronautics, Xueyuan Road 37, Beijing 100191, China
\textsuperscript{b}Deutsche Telekom Laboratories and TU-Berlin, Berlin, Germany
\textsuperscript{c}Department of Computer Science, University of Massachusetts, Amherst, United States

\textbf{A R T I C L E  I N F O}

\textbf{Article history:}
Received 9 April 2009
Received in revised form 19 December 2009
Accepted 1 March 2010
Available online xxxx

\textbf{Keywords:}
Location privacy
Anti-localization
Anonymous routing
Delay Tolerant Network

\textbf{A B S T R A C T}

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.

\( ^\text{\textcopyright} 2010\) Elsevier B.V. All rights reserved.

\textbf{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 [3–5]. This kind of routing is referred to as \textit{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]. An active area of research on routing in DTN focuses on minimizing the number of replicas while keeping delivery ratio high [8].

A message may cause others interests so they want to know the location of the author sending the message. We define this kind of messages as sensitive messages. This paper considers the adversarial localization as the chief threat to the security of nodes in 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 autocratic government may eavesdrop on the open-air wireless communication to detect those messages and localize the senders. The nature of shared transmission media makes wireless networks very vulnerable to security threats. In an un-trusted network, the author does not want adversaries to know his/her position for the reason of personal security. The motivation of this paper is to provide location anonymity for users in un-trusted DTNs.

\* Corresponding author. Tel.: +86 13520096976.
E-mail addresses: luxf@cse.buaa.edu.cn (X. Lu), Pan.hui@telekom.de (P. Hui), towsley@cs.umass.edu (D. Towsley), pujh@buaa.edu.cn (J. Pu), xiongz@buaa.edu.cn (Z. Xiong).

1389-1286/ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.comnet.2010.03.002

Please cite this article in press as: X. Lu et al., Anti-localization anonymous routing for Delay Tolerant Network, Comput. Netw. (2010), doi:10.1016/j.comnet.2010.03.002
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 at least \( n \) different receivers. ALAR is designed to meet the following objectives:

1. Minimize the probability of a sender being localized by adversaries, \( P_s \).
2. Maximize the probability that the destination receives the message after time \( t \), called \( P_r(t) \). To achieve this objective, the sender wants the message to reach as many nodes as possible. So \( P_r(t) \) is equal to the ratio of receivers to all nodes after time \( t \).

Many protocols have been proposed to provide the user location privacy: an user cannot know another user’s location from his/her connection request by the protocols [15,16]. However, if the receivers localize the source of the connection request 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 [9]. 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 [10]. 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.

ANOGR is an anonymous protocol based on on-demand routing that provides route anonymity and location privacy to MANETs [12]. In terms of route anonymity, ANOGR prevents adversaries from tracing a packet flow back to its source or destination; in terms of location privacy, ANOGR ensures that adversaries cannot discover the real identities of local transmitters. However, the privacy ANOGR provides to the sender is the identity privacy instead of location privacy.

Some secure routing protocols, such as SEAD [13] and ARAN [14], 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 [11]. Instead of encrypting the whole packet, they suggest encrypting a small piece of information about the source and destination and sending it together with the data packet. A relay node only needs to verify the small piece of information, rather than the whole packet. In ASR, Zhu’s solution makes use of shared secrets between any two consecutive nodes. The objective 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 identity 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 while it does not have a GPS device, it can employ the localization algorithm to get its location from beacon nodes which 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 [17]. 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 [18–20], 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 Fig. 1(a) shows. In Fig. 1(a), both A and B receive the sender’s signals and know signal arrival angles, 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 three beacon nodes to compute an unknown node’s location. TDOA is a widely used range-based triangle localization algorithm. 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 $t_i (i = 1, 2, \ldots, h)$.
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, $\Delta r_{ij} = (t_i - t_j) \cdot C, (i, j = 1, 2, \ldots, h, i \neq j)$, where $C$ is the speed of electromagnetic wave and $(i, j)$ is an enumeration of all pairs of receivers.
4. With two receivers, they can get a curve on which any point has the same $\Delta r$ to the sender.
5. At least three receivers with known positions are required to find a 2D-position from two TDOAs as Fig. 1(b) shows.

The main challenge of TDOA algorithm is clock synchronization. Actually, clock synchronization is not necessary to the senders in ALAR. Instead, it is necessary to the adversaries. Nodes can achieve clock synchronization by using GPS. The precision of GPS clock can be less than 10 ns which is enough to TDOA. In fact, the principle of TDOA and other range-based triangle localization algorithms are the same.

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. This kind of passive threat is hard to be detected, so it is another vital threat to MANETs [9].

It needs to be emphasized that a node’s communication range depends on its transmission power, radio propagation, antenna gain and the environment, etc. If the distance from the receiver to the sender is very long, the received transmission power at the receiver’s antenna is too weak to be distinguished from the noise because of the path loss and shadowing [23]. Therefore, we assume the node’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 attack if they know the transmitter’s real location. We assume the adversary’s detector is equipped with an omni-directional antenna and executes TDOA algorithm to localize the transmitter. Actually, if adversaries employ other range-based triangle localization algorithms, ALAR is valid also. 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 two 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 does 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 packets. As an adversary is assumed to be mobile in the network, its radio device is supposed to be powered by batteries. If an adversary localizes 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 localizes 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 two 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. An adversary does not help the sender and relays to forward packets but it exchanges segments with other adversaries.

5. Anti-localization anonymous routing

5.1. Model and assumptions

In this section, 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. When the sender wants to broadcast a packet, it sets the source and destination addresses of this packet as Broadcast Address. Relays do not modify the source and destination addresses of a packet. Therefore, the adversaries are not able to know the sender’s identity from a packet. A node broadcasts a heartbeat signal periodically which includes its identity. A node knows that it has a neighbor and its identity when it receives a heartbeat signal from another node. A node collects heartbeat signals and regards the source nodes of these heartbeat signals as its neighbors. It sends a segment only when the following two conditions are met: (1) the number of heartbeat signals received 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. If the two conditions are met and the transmission media is free, the transmitter waits for a heartbeat period plus a random time and sends the packet.

- Energy Constrained: As is introduced in the precious section, a node’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 may 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 directly in one packet. The basic idea of ALAR is to split the original message into k segments and send each segment to at least 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 message is divided into two packets. Adversary A received packet1 from nodei, nodei forwarded packet1 twice, respectively, at position 1 and position 2. Adversary A and B received packet2 from nodei, at position 1, and adversary C received packet2 from nodei 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 packeti from two 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 indicated in the last segment.
3. The sender sends each segment when (1) it has at least n neighbors and (2) the receivers of S; do not overlap with the receivers of Sj, i ≠ j.
4. After receiving a segment from others, a relay forwards the segment to its neighbors when (1) it has at least n neighbors and (2) the neighbors do not overlap with the receivers of previous segments. A relay may forward a segment multiple times, and each time it forwards the segment which has been forwarded the least.
### 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 $\lambda$. Now, we calculate $P_l$ and $P_r(t)$, respectively. First, we assume that a sender sends a sensitive message by epidemic routing.

According to the 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 Eq. (1) according to Spatial Poisson theory [24].

$$P_l = 1 - P(\text{Not being localized}) = 1 - P(||l\pi r^2|| < 3)$$

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

(2)

According to the research in [25], the pairwise meeting time between nodes is nearly exponentially distributed, if nodes move in a limited region according to common mobility models and if their transmission range is largely smaller than the length of network area, say $r \ll L$. The authors also derived the following estimation of the pairwise meeting rate $p$:

$$p \approx \frac{2 \omega r E[V^r]}{L^2}$$

(3)

where $\omega$ is a constant which depends on the mobility model used and $E[V^r]$ is the average relative speed between two nodes, $\omega \approx 1$ when the mobility model is random direction and $\omega \approx 1.3683$ when the mobility model is random waypoint. Assume the average speed of ordinary nodes is $v_1$, the average relative speed $E[V^r]$ can be calculated by

$$E[V^r] = \frac{\omega v_1}{\pi r^2} \int_0^{2\pi r} \frac{x^2}{\sqrt{1 - \left(\frac{x}{2\pi r}\right)^2}} dx$$

Eq. (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 duration from the beginning of packets sending till present [4].
The segment received by node $i_k$ at time $t_k$ is equal to the ratio of receivers to all nodes after time $t$, so it can be got by

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

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

$$CP(t) = \left(1 - P_i\right) \times P_i(t) = \sum_{j=0}^{\lfloor \frac{k}{r} \rfloor} \frac{e^{-\pi x^2 j r^2}}{1 + e^{-pN(N-1)}} \tag{6}$$

5.3.2. ALAR

By ALAR, the transmitter cannot send a segment as casually as by epidemic routing. Let the probability that a node has at least $n$ neighbors be $\varepsilon$.

$$\varepsilon = P(|\pi r^2| \geq n) = P(|\pi r^2| > 0) - 1 - P(|\pi r^2| = 0). \tag{7}$$

As $P(|\pi r^2| \geq n) > 0$, $P(|\pi r^2| > 0) = 1 - P(|\pi r^2| = 0)$. Then, Eq. (7) can be transformed into

$$\varepsilon = \frac{P(|\pi r^2| \geq n)}{1 - P(|\pi r^2| = 0)} = \frac{1}{1 - x} \cdot x = \frac{\pi r^2 (N + M)}{L^2} \tag{8}$$

We can know that the relays have fewer opportunities to forward a segment by ALAR comparing with by epidemic routing. Let the ratio that receivers can receive a segment be $P^{(i)}(t)$ in ALAR and the ratio that they receive $k$ segments be $P^{(k)}(t)$.

$$P^{(i)}(t) = \frac{1}{1 + e^{-pN(N-1)}} \tag{9}$$

$$P^{(k)}(t) = P^{(i)}(t)^k = \left(\frac{1}{1 + e^{-pN(N-1)}}\right)^k \tag{10}$$

When the sender sends $S_k$, if there are more than two adversaries within its transmission range and at least one of them has received $k - 1$ segments that the sender sent, these adversaries are able to know the content of the 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 adversaries can receive segments both from other adversaries and ordinary nodes. Thus, the total nodes in the area that can receive messages is $(N + M)$. 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

$$I_i(t) = \frac{(M + N)}{1 + e^{-pN(M+M-i)(N+N-1)}(M + N - 1)} \tag{11}$$

As time increases, the number of nodes received segments increases accordingly. When the sender sends $S_i$ at time $t$, $I_{k-1}(t)$ is supposed to be smaller than $I_{k-2}(t), \ldots, I_{1}(t)$. So, we employ $I_{k-1}(t)$ as the approximate number of nodes that have received segments $S_1, S_2, \ldots, S_{k-1}$. Let the time interval between $t_k$ and $t_{k-1}$ be $T_k = t_k - t_{k-1}$ and the area which $S_{k-1}$ can be carried to be $A$. The maximum of $A$ is $L^2$.

The density of adversary nodes that have received the segment $S_k$ at time $t_k$ is $A$. Then, $I_{k-1}(t_k) M / A$ is the probability that receivers can receive a segment by ALAR comparing with by epidemic routing. We can know that the relays have fewer opportunities to forward a segment by ALAR comparing with by epidemic routing.

$$A = I_{k-1}(t_k) M / (M + N) A = (1 + e^{-pN(M+M-i)(M + N - 1)}) \tag{12}$$

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, $S_1, \ldots, S_{k-1}$.

$$P^{(k)}(t) = \left(1 - \frac{1}{\lfloor \frac{k}{r} \rfloor} \frac{e^{-\pi x^2 j r^2}}{l!}\right) (1 - e^{-\pi x^2}) \tag{13}$$

$$A = \frac{M}{(1 + e^{-pN(M+M-i)(M + N - 1)})} \tag{14}$$

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

5.4. Evaluation

In this section, we compare $P_i$, $P_i(t)$ 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 $P_i$

Fig. 2 shows that with the increase of $M$. $P_i$ of epidemic routing increases sharply to 0.83 and $P_i$ 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%.

Fig. 3 shows the impact of $n$ and interval on $P_i$. The probability of there being more than two adversaries within the sender’s transmission range is low when $n$ is very small. So the probability $P_i$ is low correspondingly. In this case, the average number of neighbors is nine. In Fig. 3(a), when the value of $n$ is near to the average number of neighbors

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>200</td>
<td>$M$</td>
<td>1, 100</td>
</tr>
<tr>
<td>$r$</td>
<td>120 m</td>
<td>$L$</td>
<td>1000 m</td>
</tr>
<tr>
<td>$v$</td>
<td>60 m/min</td>
<td>$v_i$</td>
<td>60 m/min</td>
</tr>
<tr>
<td>$k$</td>
<td>2</td>
<td>Interval</td>
<td>10 min</td>
</tr>
</tbody>
</table>

Please cite this article in press as: X. Lu et al., Anti-localization anonymous routing for Delay Tolerant Network, Comput. Netw. (2010), doi:10.1016/j.comnet.2010.03.002
of a relay, \( P_l \) increases rapidly. Besides, a relay has few chances to have lots of neighbors at one moment, so it has few opportunities to send segments and the probability \( P_l \) is very low when \( n \) is large.

Fig. 3(b) shows the impact of the time interval on \( P_l \). The adversaries which receive \( S_{k-1} \) from the sender are not able to move far away during a short time, so they have a high 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 \( T \) increasing from 1 to 8 min, \( P_l \) decreases by over 97%. \( P_l \) is the 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 \( S_{k-1} \) during that interval and \( A \) 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 sender will be localized with the probability 100% because the receivers can receive its segments, wherever the sender moves.

5.4.2. The study of \( P_r(t) \)

Fig. 4(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 \( P_r(t) \). It indicates that 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. Fig. 4(b) shows that the larger the \( n \) 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 because 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 \( n = 3 \). If \( n \) is very large, say \( n > 15 \) in this case, the delivery process is extremely slow as this figure shows.

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. Fig. 5(a) shows \( CP(t) \) of ALAR is smaller than that of epidemic routing in the beginning phase because \( CP(t) \) of ALAR is much smaller than that of epidemic routing in that phase as Fig. 4(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 three segments than to receive two segments.

Fig. 5(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 Fig. 4(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 did above, 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 DTNs 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 if others read it after some days. Here, if we set $n = 1$, the delivery delay would reduce because relays have more opportunities to forward a segment, 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 Eq. (10) that with the increase of $k$, it takes longer time for $P_t(t)$ to achieve 1. So we suggest $k = 2$ for less delivery delay. From the study of the impact of $n$ on $P_t$, we know that $P_t$ decreases with the increase of $n$ when $n$ is larger than a threshold, on the contrary, the delivery delay increases. Therefore, we suggest the appropriate value of $n$ to be from 2 to 4.

In our current study, we do not consider the energy consumption and we assume single point communication, which is not in favor for $P_t(t)$, but can act as a lower bound for the performance. We think the proportion of sensitive messages vs. all messages does affect the anonymity performance. If all messages are sensitive, adversaries will localize any sender without checking the sensitivity of messages. Even ALAR cannot have good anonymity performance because the precondition of ALAR vanishes. On the contrary, if most of messages are unsensitive, adversaries will balance the energy consumption of calculations with the veracity of localization. Also, we think the probability distribution of the sensitive messages also impacts the anonymity performance. If most sensitive messages are launched from somewhere, adversaries may go to that place so the density of adversaries at that place is much higher than that at other places. This will lower the anonymity performance of ALAR at that place.
6. Validation

As human mobility plays a key role in packet delivery in DTN [27], we checked the user contacts in the real world. We validated ALAR on two real-world experiment datasets to determine the impacts of human mobility on the routing of packets.

6.1. Contact 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 iMotes sent heartbeat signals every 2 min. The format of the 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, \(\text{contact duration} = t_e - t_b\).

Fig. 6(a) shows the distribution of the number of neighbors. We conclude that each node has at least one neighbor with around 30% experiment time 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 six neighboring nodes at one moment in this experiment.

Fig. 6(b) shows the distribution of the contact durations. The statistical study of these contacts shows that more than 80% of the contacts are shorter than 10 min and more than 90% contacts are shorter than 20 min. 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 \(n\), we set \(n = 4\) in the simulation. Also, we perform routing simulations with different \(n\) on the Infocom dataset to study the impact of \(n\). We compare \(P_t\), \(P_t(t)\) and \(CP(t)\) of ALAR with that of epidemic routing. To get the average \(P_t\), \(P_t(t)\) and \(CP(t)\), we run the simulation program 1000 times with each \(n\). In each simulation, we randomly select about 12.5%, 25%, 37.5% and 50% nodes as the adversaries.

6.2.2. Anonymity performance: \(P_t\)

Fig. 7(a) shows the impact of \(n\) on the sender’s \(P_t\). When \(n\) is lower than 4, \(P_t\) of ALAR is lower than 0.03 even when the ratio of adversary to all nodes is about 50%. However, \(P_t\) increases sharply when \(n\) is 4 or 5. The reason is that if the sender sends a segment only when it has at least 1–3 neighbors, the probability of there being at least three adversaries in its communication range is small. If the sender sends a segment when it has more than three neighbors, the probability of there being more than two adversaries within its communication range increases. However, when \(n\) is larger than 5, \(P_t\) decreases to 0. As Fig. 6(a) indicates, a node has few chances to have more than 6 neighboring nodes at one moment, so the sender has almost zero opportunity to send segments when \(n \geq 6\). Therefore, \(P_t\) is zero when \(n \geq 6\).

In Fig. 7(a), \(P_t\) of ALAR is largest when \(n = 5\). That means the worst anonymity performance of ALAR. Therefore, we compare the worst anonymity performance of ALAR with the normal anonymity performance of epidemic routing. Fig. 7(b) shows that with the increase of the number of adversaries, both \(P_t\) of ALAR and epidemic routing increase correspondingly. The increasing rate of \(P_t\) of epidemic routing is much faster than that of ALAR. If there are 20 adversaries in the network, the sender’s probability of being localized is larger than 0.8 if it employs epidemic routing, but \(P_t\) of ALAR is just about 0.15. This statistical study also validates that ALAR can increase the sender’s anonymity performance by over 81% with 50% networks nodes being adversaries.

6.2.3. Delivery ratio: \(P_t(t)\)

Fig. 8 shows the packet delivery process by ALAR and epidemic routing. The increasing of \(P_t(t)\) of ALAR when
n < 3 is close to that of epidemic routing. When n < 5, P_r(t) of ALAR after 1 day is larger than 0.95. P_r(t) of epidemic routing after 1 day is 1. It shows ALAR may cause 5% reduction in the delivery ratio. With the increase of n, the delivery delay increases as well. It certifies the correctness of the P_r(t) model of ALAR.

There are four durations during which P_r(t) remains almost unchangeable. The four durations can be clearly found from the curve of n = 5 in Fig. 8. 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)

Fig. 9 shows the comparison of CP(t) of ALAR with that of epidemic routing after 1440 min under 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 small, a node has few opportunities to have more than three adversaries within its transmission range. Therefore, the sender’s probability of being localized is very small even if it uses epidemic routing. Thus, the advantage of ALAR is not clear. However, with the increase of the number of adversaries, CP(t) of epidemic routing decreases rapidly and becomes much lower than that of ALAR. When the ratio of adversaries to all nodes is 50%, CP(t) of epidemic routing after 1 day 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 experiment dataset over 30 min and found the average number of neighbors is 6.2. We assign \( n = 4 \) and \( k = 2 \) in this case. We also checked \( P_s, P_s(t) \) and \( CP(t) \) when \( n = 2 \) and \( n = 7 \). The simulation results are listed in Table 4. It shows that \( CP(t) \) achieved the largest value when \( n = 4 \). It verifies that the conclusion about the value of \( n \) in Section 5.5 is reasonable.

Table 4

<table>
<thead>
<tr>
<th>( k )</th>
<th>( n )</th>
<th>( P_s )</th>
<th>( P_s(t) )</th>
<th>( CP(t) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>0.027</td>
<td>0.45</td>
<td>0.438</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.038</td>
<td>0.64</td>
<td>0.616</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>0.048</td>
<td>0.37</td>
<td>0.352</td>
</tr>
</tbody>
</table>

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 datasets to certify the advantage of ALAR. The validation simulations show that ALAR can decrease the sender's probability of being localized by at least 81% with about 5% loss in \( P_s \). 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. We did not consider the energy consumption, the proportion of sensitive messages and the distribution of the sensitive messages in our current study. In the future, we will consider these issues and study how to assure high anonymity under different conditions.

Acknowledgments

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

References


Pan Hui is currently a senior research scientist in the Deutsche Telekom Laboratories and TU-Berlin after 1 year of postdoc in Cambridge on SOCIALNETs, ITA related projects. He received his B.S. and MPhil Degrees in 2001 and 2003 in computer science from University of Hong Kong. He received his Ph.D. in computer science from university in 2007. His research interests include social networks and Delay Tolerant Networks.

Xiaofeng Lu is currently a Ph.D. candidate in the School of Computer Science and Engineering, Beijing University of Aeronautics and Astronautics. He received his B.S. and M.S. degrees in computer science from North China University of Technology, in 1999 and 2005, respectively. He visited the Computer Laboratory of University of Cambridge from 2007 to 2008. His research interests include wireless communication, network security, and anonymous communication, etc.

Pan Hui

Don Towsley received the B.A. degree in physics and the Ph.D. degree in computer science, both from University of Texas University, in 1971 and 1975, respectively. He is currently a Distinguished University Professor in the Department of Computer Science at the University of Massachusetts – Amherst, where he co-directs the Networking Research Laboratory. He has been a Visiting Scientist at AT&T Labs – Research, IBM Research, INRIA, Microsoft Research Cambridge, and the University of Paris. He currently serves as Editor-in-Chief of the IEEE/ACM Transactions on Networking and on the editorial boards of Journal of the ACM and IEEE Journal of Selected Areas in Communications. He has twice received IBM Faculty Fellowship Awards, and is a Fellow of the IEEE and the ACM. Dr. Towsley’s research interests include network measurement, modeling, and analysis.

Xiaofeng Lu

Juhua Pu is currently an Associate Professor in the School of Computer Science and Engineer of Beijing University of Aeronautics and Astronautics. She received her Ph.D. in computer science from Beijing University of Aeronautics and Astronautics in 2005. In 2004, Juhua Pu visited the Hong Kong University of Science and Technology. Dr. Pu’s current research interests include sensor networks and distributed multimedia, etc.

Don Towsley

Zhang Xiong is currently a Professor in the School of Computer Science and Engineer of Beijing University of Aeronautics and Astronautics. He received his M.S. degree in computer science from Beijing University of Aeronautics and Astronautics in 1984, and visited the Michigan State University from 1989 to 1992. Professor Xiong was awarded China National Golden Medal for Progress in Science and Technology in 1994. His current research interests include sensor networks, multimedia, etc.

Juhua Pu

Zhang Xiong