Anonymous Geo-Forwarding in MANETs through Location Cloaking

Xiaoxin Wu, Jun Liu, Xiaoyan Hong, and Elisa Bertino, *Fellow*, *IEEE*

Abstract—In this paper, we address the problem of destination anonymity for applications in mobile ad hoc networks where geographic information is ready for use in both ad hoc routing and Internet services. Geographic forwarding becomes a lightweight routing protocol in favor of the scenarios. Traditionally, the anonymity of an entity of interest can be achieved by hiding it among a group of other entities with similar characteristics, i.e., an anonymity set. In mobile ad hoc networks, generating and maintaining an anonymity set for any ad hoc node is challenging because of the node mobility and, consequently, the dynamic network topology. We propose protocols that use the destination position to generate a geographic area called an *anonymity zone (AZ)*. A packet for a destination is delivered to all the nodes in the AZ, which make up the anonymity set. The size of the anonymity set may decrease, because nodes are mobile, yet the corresponding anonymity set management is simple. We design techniques to further improve node anonymity and reduce communication overhead. We use analysis and extensive simulation to study the node anonymity and routing performance and to determine the parameters that most impact the anonymity level that can be achieved by our protocol.

Index Terms—Ad hoc routing protocol, anonymous routing protocol, georouting protocol, anonymity, communication privacy.

1 Introduction

Mobile ad hoc networks (MANETs) are envisioned as an effective solution for extending infrastructure-based wireless network communications and/or self constructing when fixed infrastructures are not available. For example, wireless mesh networks and WiMax both include a MANET as an optional extension to the multihop wireless infrastructure. First, responder networks and vehicular networks also take much of their form as MANETs [1], and MANET has been used in tactical applications. Communication privacy in MANETs is of increasing concern for a large variety of application domains, and therefore, techniques for achieving high privacy assurance are required. An important privacy requirement for MANETs is represented by the anonymity of the communicating parties. In general wireless networks, many scenarios have shown that anonymity is critical. For example, the relationship of the identities of WLAN or cellular users and their locations needs to be hidden from third parties [2], [3], locations of sources in sensor networks should not be traced by malicious nodes [4], and active paths and network topology need to be protected in MANETs; otherwise, nodes could be traced [5].

In this paper, we investigate an application scenario under which Location-Based Services (LBSs) [6], [7] are accessed from a MANET segment. Anonymity is particularly crucial when an ad hoc node receives sensitive data from well-known servers. This receiver may not wish its identity to be revealed to the network or any third party while communicating with those servers. We refer to this requirement as *destination anonymity*. In this work, we use client/destination and server/source interchangeably.

In this application scenario, as geographic information is required for LBSs, it is then natural to use geographic-position-assisted routing protocols to deliver data from the server to the client [8], [9], [10]. Such a routing protocol has the advantage of being scalable and lightweight. In addition, geographic-position-assisted routing has the potential of supporting identity anonymity in that a node's position, not its ID, is required in routing. For example, in Greedy Perimeter Stateless Routing (GPSR) [10], node positions are used; thus, the real identity of a node, e.g., a destination, is hidden. The local and stateless routing operation also helps preserve privacy. However, in many applications [13], [12], position information is sensitive data. Presenting the position during the communication gives the adversary the opportunity to trace down a destination according to its position and then identify the destination in a face-to-face manner. Through such a *probing attack*, adversaries can easily link nodes' locations to their identities.

The goal of our paper is to explore the advantages of geographic-assisted routing while, at the same time, dealing with the probing attack mentioned above, i.e., to address the privacy problem when using the aforementioned sensitive position data. We will focus on achieving destination anonymity in this paper. When the servers are well known, source anonymity is less important. On the other hand, if source anonymity is required to achieve

- *X. Wu is with the Intel Communication Beijing Lab, Raycom Infotech Park A, Floor 8, No. 2 Kexueyuan South Road, Haidian, Beijing, China, 100080. E-mail: xiaoxin.wu@intel.com.*
- *J. Liu and X. Hong are with the Department of Computer Science, University of Alabama, Tuscaloosa, AL 35487-0290. E-mail: {jliu, xyh}@cs.ua.edu.*
- *E. Bertino is with the Department of Computer Sciences, Purdue University, 250 N. University Street, West Lafayette, IN 47907-2066. E-mail: bertino@cerias.purdue.edu.*

Manuscript received 15 May 2007; revised 10 Oct. 2007; accepted 21 Jan. 2008; published online 11 Feb. 2008.
Recommended for acceptance by S. Das.
For information on obtaining reprints of this article, please send e-mail to: tpds@computer.org, and reference IEEECS Log Number TPDS-2007-05-0154.
Digital Object Identifier no. 10.1109/TPDS.2008.28.

1045-9219/08/$25.00 © 2008 IEEE Published by the IEEE Computer Society

unlinkability between a client and the server, a packet can be transmitted without carrying source identity.

We propose an anonymous geographic routing algorithm that uses fuzzy destination positions. The notion of fuzzy position is used in privacy-preserving LBSs [13], [14]. Under such an approach, a mobile user intentionally provides inaccurate positions to services in order to protect its real positions. Here, we use a fuzzy position in geographic forwarding (geo-forwarding) to prevent adversaries from obtaining the real position of a node and, therefore, to prevent a destination ID from being discovered based on its position. A client generates a pseudo destination (PD), which is chosen in such a way that when data packets are forwarded to the location, the real client has a high probability of receiving them. Such a position is sent to the application server, toward which the server sends packets. The successful delivery in such a routing algorithm relies on the broadcast nature of wireless communication, where a transmission can always be received by all the nodes within the transmission range of the sender. Therefore, if the real destination is located in a geographic area that is not far away from the PD or from the forwarding path, it will receive the packets. Such a geographic area, which we refer to as an *anonymity zone (AZ)*, is the key concept of our design. The destination anonymity is determined by the number of nodes that are located in the AZ, and the protocol is thus called the *zone-based anonymous positioning routing (ZAP) protocol*.

The main design challenges of ZAP are how large anonymity sets are maintained and how data packets, given only pseudo location information, are delivered. The major contributions of this paper are two approaches tackling the problems, namely, a destination-based AZ and a route-based AZ. The former uses a geographic area around the fuzzy destination as the AZ. Based on the concept, a basic ZAP protocol, namely, ZAP with PD (PD-ZAP), and dynamically expansible zone, namely, geocasting-based anonymous protocol (G-ZAP), are proposed. The latter approach turns the entire routing path with extra hops into an AZ, i.e., ZAP with Route Redundancy (RR-ZAP).

ZAP is a best effort protocol. It provides destination privacy, measured by the size of the anonymity set, with probability. While a guaranteed anonymity level is not a design goal, G-ZAP and RR-ZAP provide effective methods for increasing the anonymity set size. We present analysis showing how the anonymity set is affected by nodal density, mobility, and communication durations. In reality, a node can estimate local nodal density and mobility through analyzing neighborhood transmissions. Compared to the approaches that achieve a predefined anonymity size, e.g., multicasting to a fixed number of nodes, the idea proposed here greatly simplifies the operation and significantly reduces the overhead, with a tolerance probabilistic anonymity level. In addition, the three variants reveal different advantages by the nature of the design, e.g., an efficient "geocast" strategy with G-ZAP, reduced "hot spots," and strategy of maximizing entropy in RR-ZAP. We develop mathematical models for analyzing different parameters in order to obtain guidelines for AZ management. Both simulation and analytical work are presented in the paper.

This work is based on our early paper [34], with substantial new and revised content on scheme design, analysis, and evaluations. The remainder of this paper is organized as follows: Section 2 introduces system and attack models that are relevant to our protocols. The section also describes the service model of the ZAP protocol. Section 3 presents the details of the ZAP protocol, introducing the three variants in sequel. Section 4 discusses protocol anonymity and the attack. Section 5 presents analysis for several key design metrics. Section 6 reports evaluation results from both the numerical analysis and the simulation. Section 7 discusses related works. Finally, Section 8 concludes this paper.

2 System and Attack Models

2.1 Network and Attacker Models

The ZAP framework supports network connection through MANETs with lightweight geo-forwarding while preserving destination anonymity. The information about the servers, including their positions, are well known. Servers' public keys can be obtained by mobile nodes before they join the network. We assume that the network is not sparse (e.g., the node density is more than 50 $\mathrm{nodes/km^2}$) and each node has an equal probability of being a receiver (client). Each node knows its own position, e.g., through a GPS system. To facilitate geo-forwarding, nodes broadcast their positions locally through "hello" messages. We further assume that the wireless channel is bidirectional, and therefore, the multihop path between any two communication ends is also bidirectional.

The attackers that we consider are ad hoc nodes located in the network, which monitor or trace the behavior of other nodes for malicious purposes through eavesdropping the communication channel or participating network functions (e.g., routing). Attackers are passive. They follow the protocols, and would function normally, in particular when included in any active routes. They do not act aggressively to interrupt the correct network function in order to obtain additional information, because they would like to stay in the network without being noticed. This assumption leaves security concerns such as black/gray holes, *man in the middle*, *denial of services*, and jamming as orthogonal problems. An attacker can collect the position information of its neighbors by intercepting hello messages. An attacker or colluding attackers can therefore discover the local network topology. This ability enables the attackers to perform a so-called *intersection attack* to violate the destination's anonymity. The details will be given in Section 4. In this work, we assume that the attackers do not have the ability to form a global monitoring network. We also assume that the attackers cannot pinpoint the location of a particular transmission. These assumptions allow us to justify the design to easier to launch attacks, like curious eavesdroppers, without loss of generality. Sophisticated techniques exist for locating a transmitter, but that will require special devices, higher cost, and scanning time. On the contrary, an attack like the one assumed by our attacker model can be easily perpetrated in many cases.

2.2 Communication Model

ZAP is designed for destination anonymity in LBSs where geographic locations can assist routing through MANETs. The basic ZAP routing protocol is a greedy geo-forwarding scheme that applies directly to data packet delivery. The greedy geo-forwarding is used in several MANET routing protocols [8], [9], [10]. In such a routing protocol, the source obtains the location of the destination in the beginning, and then, a forwarder will always select the node that is geographically closest to the destination as its next hop. The next-hop generation relies on a local information exchange (an ad hoc node exchanges its position information with its neighbors through periodic "hello" messages). Each node will thus maintain a neighbor list to record all the locations. As greedy geo-forwarding always tries to move a packet closest to the destination, the resulting route generally has a hop count that is approximately minimum, while having a small hop count is a very desired feature in multihop networks for throughput enhancement and routing management simplification. In what follows, the service model and the message format are given.

The application starts with a client (*destination*) that sends a server (*source*) a connection request to initiate the session. The request indicates the protocol-specific parameters for setting up a private route, e.g., the fuzzy position information and, if necessary, the range of the AZ. The range of the AZ is determined by the client's desired anonymity level and its local node density that can be estimated by receiving "hello" messages from its neighbors. The connection request can be sent through greedy geographic forwarding, by traditional routing algorithms, or by flooding to the server. To assure data confidentiality and integrity, the destination generates a symmetric key and sends it to the server as well. The key and the client-specific parameters are encrypted by the server's public key and are carried in the connection request. HMAC [16] is used to protect the integrity of the request message, using the symmetric key carried in the message.

The source retrieves the *position of the PD* field after receiving the connection request. It then initiates the greedy geo-forwarding to deliver data packets to the PD. Using greedy geo-forwarding, any forwarder (including the source) forwards the data packet to a neighboring node that is closest to the PD. The data will finally be transmitted in the AZ where the destination is located and will be received by the destination. The source uses the symmetric key to encrypt data and uses HMAC for data integrity.

The message frames for connection requests and data packets can be structured as shown in Fig. 1. The contents of the *Routing information* field in a connection request depends on the routing algorithm used for sending the request. If geo-forwarding is used, the field will be the server's geographic location. For data packets, the PD location information can provide the attacker chances to trace the packet flow to the vicinity of the real destination. This attack and related mitigation techniques are discussed in more detail in Section 4. The *Next-Hop ID* field becomes a broadcast address when the forwarder is a proxy, which does not reveal additional information.

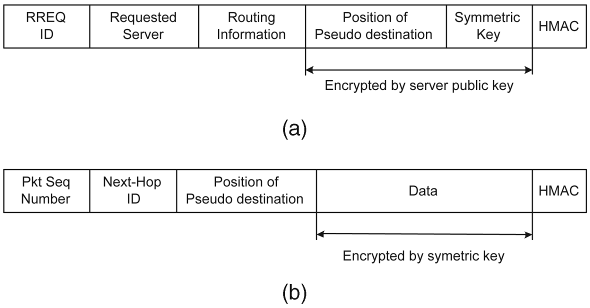

RREQ ID
Requested Server
Routing Information
Position of Pseudo destination
Symmetric Key
HMAC
Encrypted by server public key
(a)
Pkt Seq Number
Next-Hop ID
Position of Pseudo destination
Data
HMAC
Encrypted by symetric key
(b)

Fig. 1. Packet frame structure.

Concerning breaching the client anonymity through this request message, our claim is that the probability for an attacker of intercepting a request and determining its initiating location is very small. This is due to a few reasons. First, the identity of the request originator is not carried in the message (see Fig. 1). Second, the same request messages will be forwarded by many intermediate nodes in addition to the originator. Next, the generation of a request by a client is sporadic. Last, fuzzy position information in a packet does not help much in pinpointing the transmitting node.

ZAP does not apply traditional packet-based ACKs, as a continuous ACK streaming sent out from the destination may lead attackers to trace their originator. In our approach, the destination sends an ACK for the first received packet. After that, it sends NACK if any packets are missing. The ACK and NACKs are sent along independent paths based on the most current locations of the source and the real destination. In order to preserve traditional Internet application semantics, such a NACK-based acknowledgment mechanism needs middleware support at the server side. The middleware hijacks the connection and mimics consecutive ACKs as required by the TCP protocol in the absence of real ACKs. When a NACK is received, the middleware will trigger a retransmission by either generating three duplicate ACKs or sending a delayed ACK to force a time-out.

A new session has to be started if the destination can no longer receive data packets, typically when the destination has moved away from the AZ. The destination will be aware of such a situation when it cannot receive a missing packet, even if it has sent a few NACKs (the maximum number of NACKs that can be sent for a packet is a system parameter). In this case, the destination has to send a new connection request, along with the updated PD information. Based on the information, the source initiates another private route.

3 Zone-Based Anonymous Positioning Routing Protocol

The ZAP protocol preserves destination anonymity through the use of AZ, under which a destination is colocated with a number of other nodes. The key idea is to create an *AZ* based on a carefully selected PD. The following sections present different methods in selecting the PD and creating the AZ. Mainly, the ZAP protocol has two distinct variants: one uses destination-based AZ, and the other uses route-based AZ. Anonymity analysis and

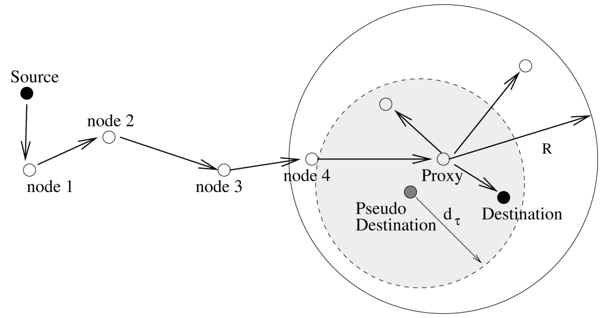


Fig. 2. The PD-ZAP approach.

performance analysis on these methods are given in later sections.

3.1 ZAP with Pseudo Destination (PD-ZAP): The Basic Approach

The basic ZAP, referred to as PD-ZAP, uses a PD and an AZ around it (see Fig. 2). The real destination generates a *PD*, which has a random position not too far from that of the real destination. The PD's position is carried in request and data packets in the PD field. Therefore, the connection request does not have to carry the real identity of the destination, as it is not required for routing. This guarantees the destination anonymity, even if the source is compromised.

In PD-ZAP, a packet will finally be received by a node that is closest to the PD.[1] This node then acts as a *proxy* and broadcasts the received packet to all of its neighbors. In this paper, a broadcast is defined as the process according to which a node transmits a message to all its neighboring nodes that are within its radio coverage. In Fig. 2, the solid circular represents the transmission range of the proxy, which has a radius of R. R is the maximum ad hoc channel coverage. If the real destination is within the proxy's radio coverage, it will receive the data packet.

The generation of the PD is the key part of the algorithm. The maximum distance (or the distance threshold value) between the PD and the real destination, denoted as d_τ, determines both the node anonymity and the success of a packet delivery. The distance cannot be too long; otherwise, the real destination may not receive the data packet from the proxy. It cannot be too short either, because a small distance results in a small anonymity set. As shown in Fig. 2, the destination AZ (D-AZ) in PD-ZAP is the shaded circular area that is centered at the PD and has a radius of d_τ. To attackers, only one node located in that area can be the destination. The PD selection depends on node density and node mobility. The impact of the distance threshold value on anonymity set and packet delivery failure is further investigated in Section 5.2.

In PD-ZAP, the position of the PD is also used as the session ID, according to which a node receiving a packet from the proxy knows whether it is the destination. For nodes that are within the proxy's broadcast range, only the destination will be able to decrypt the packet using the established symmetric key. Other nodes simply drop the packet. During the same session, proxies can be different. This is because when different packets arrive at the D-AZ, the node that is closest to the PD may be different.

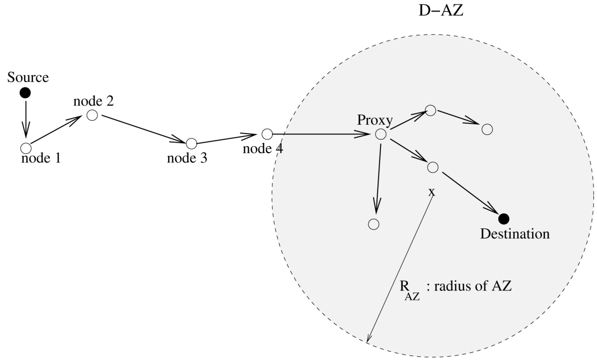


Fig. 3. The G-ZAP approach.

3.2 G-ZAP: Geocasting Anonymous Approach

In PD-ZAP, the distance between the PD and the real destination cannot be too large. Thus, the anonymity set cannot be large, especially when the node density is low. To address such a problem, we propose G-ZAP, which uses a relatively large D-AZ for improving destination anonymity. As the destination may not be able to receive the packet directly from the proxy, an approach similar to Geocasting [17] is applied, based on which a packet is locally flooded in a geographic area.

In G-ZAP, a destination selects a circular area within which it is located as its D-AZ. The source then sends packets toward the center of this area. Information about the D-AZ, such as the position of the center and the radius, is carried by data packets for routing purposes. Located in the D-AZ, any node that is the first one to receive the data packet becomes a proxy and floods the packet within the entire area. The proxy may not be the node that is closest to the center of D-AZ. As flooding is generally robust, the destination can receive the packet, as long as it stays in the D-AZ.

The general idea of G-ZAP is illustrated in Fig. 3, where R_{AZ} is the radius of the D-AZ. Note that R_{AZ} can be much larger than d_τ in the PD-ZAP approach.

Reducing the communication overhead is necessary in the G-ZAP design. G-ZAP achieves a higher destination anonymity by using a larger D-AZ, thus increasing the number of nodes residing in the D-AZ. When using a traditional flooding algorithm, every node in the D-AZ will rebroadcast the data packet. Such an algorithm results in a large overhead, and the overall network performance such as end-to-end throughput or delay may degrade significantly.

Advanced flooding algorithms that reduce the redundant broadcasts while guaranteeing message delivery have been widely investigated in MANETs and sensor networks [20], [18], [19]. These algorithms can be applied to ZAP. Since in ZAP, a flooding is processed in a known area and each node knows its position, the region-based flooding algorithm proposed in [21] can also be used to reduce the communication overhead.

When using region-based flooding, a D-AZ is further divided into a number of hexagon-shaped subregions. The

[1] It is not necessary that a node is located at that position.

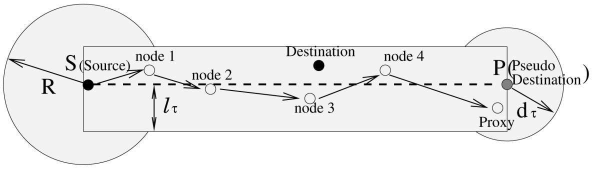


Fig. 4. The RR-ZAP approach.

length of the diagonal of each hexagon must not be greater than the maximum ad hoc radio coverage. By constructing the subregions according to such an approach, a transmission from a node in a subregion can be received by all the other nodes in the same subregion. The transmission can also be received by nodes in some of their neighboring subregions, through which the packet can be broadcast in these subregions as well.

The number of subregions depends on the size of D-AZ. The number of transmissions required for distributing a packet within the D-AZ is bounded by the number of subregions. For details about the region-based flooding algorithm, we refer readers to [21].

3.3 RR-ZAP: ZAP with Route Redundancy

Both PD-ZAP and G-ZAP build an AZ around the destination. This could generate a trade-off problem between the AZ size and the communication overhead within a hot region. Here, we present a strategy that creates an AZ along the forwarding path. This scheme makes it possible to obtain a larger AZ without incurring heavy communication load in a spot.

We propose to use a route with redundant hops to increase the size of AZ. We call this approach *RR-ZAP* (refer to Fig. 4). Like PD-ZAP, in RR-ZAP, a client (destination) creates a PD, denoted by P in the figure, to build a private route. Data packets are delivered toward the PD. They will finally be received by a proxy and are locally broadcast.

Unlike PD-ZAP, in RR-ZAP, P is not close to the real destination but can be a few hops away. P is selected so that the distance between the real destination and the direct connection between the source and the PD, which is the dashed line SP in the figure, is below a threshold value l_τ. Generally, geo-forwarding tends to use a path close to the straight line linking the source and the destination SP, making the calculation of the PD relatively easy. If the network node density is not too low, the delivery path may not deviate too far away from line SP. As long as the real destination is close to the path, it can receive the data from a node on the route, e.g., the source, the proxy, or any forwarder. In Fig. 4, the real destination can *intercept* the packet, probably from node 3.

Other than the distance between the source and the destination, the threshold value l_τ is another parameter that determines the anonymity set. It depends on node density and distribution. To an attacker, as the destination can be any node that is not more than l_τ away from the SP, the AZ for the destination then includes the shaded rectangular area in the figure. Other than that, the real destination can also be located at the circular shaded areas at the two ends of the path, which are respectively the coverage of the source and the AZ for PD-ZAP. By design, the equivalent AZ is larger than that in PD-ZAP and can be comparable to the AZ in G-ZAP.

Compared to the routes generated by PD-ZAP, routes generated by RR-ZAP are longer due to the redundant hops. Yet, this redundancy could be smaller than that introduced by G-ZAP. In addition, the RR-ZAP does not introduce hot spots, that is, small areas that have a large number of transmissions.

When an immediate acknowledgment from the real destination to the source is required, the real destination uses the same ACK/NACK mechanism, i.e., it sends an ACK for the first received packet and NACKs for the subsequent missed packets.

In RR-ZAP, an important design issue is how far the PD should be from the source, that is, how routes with redundant hops can be generated. A straightforward approach is to always require that a RR-ZAP route has a hop count equal to the maximum hop count N_{max} that a route may have. Under this approach, the destination has the largest anonymity set but at the cost of high communication overhead and, possibly, high unsuccessful delivery rate. It may not, however, be necessary to fix the size of the anonymity set to the maximum if the user anonymity requirement is not high, while, on the other hand, the routing performance is a more important concern.

In this paper, we propose to use a random match for the redundant route generation. If, when applying PD-ZAP, a route has a hop count of N, where $1 \leq N \leq N_{max}$, and N_{max} is the maximum hop count for any route in the network, in RR-ZAP, the source can select a PD so that the generated RR-ZAP route has a hop count between $max(N_{min}, N)$ and N_{max}, where N_{min} is the minimum hop count for a RR-ZAP route that can satisfy the anonymity requirement. More details about such a random match are presented in Section 5.3.

4 Anonymity, Weaknesses, and Mitigation Techniques

In this section, we discuss the anonymity properties of the ZAP protocols. We devise possible privacy attacks and propose counter measures.

4.1 Anonymity

Destination anonymity is defined to hide a destination among a set of ad hoc nodes, formally defined as *anonymity set* [22] for the destination. The size of the anonymity set measures the degree of anonymity of the destination. The larger the set size, the stronger the protection. Under our attacker model, if an eavesdropper(s) hears that n neighbors are located in the AZ (through hello messages) but is not able to determine which one is the destination, the destination has an anonymity set of size n, i.e., it is hidden among n nodes. In ZAP, the destination anonymity depends on the size of the group formed by the nodes that are located in the AZ. The key factors determining the size are the node distribution, the size of the D-AZ, the node mobility, and the session length.

As mentioned earlier, intercepting a connection request may not help attackers in identifying the destination. As the identity of the request originator is not carried in the

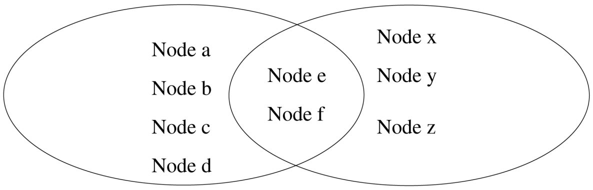


Fig. 5. Example of an intersection attack.

message, upon intercepting a request, the attacker cannot even tell whether the node from which it intercepted the message is the originator or just a forwarder.

For the same reason, in PD-ZAP and G-ZAP, when a destination sends an ACK (or NACK) back by using an alternative private route and the ACK (or NACK) is intercepted by attackers, the identity of the destination cannot be discovered. However, in RR-ZAP, by intercepting the ACK/NACK packets, the attacker may have more information on the exact hop count. In particular, when an attacker close to the RR-ZAP path never intercepts any ACK/NACK, it can estimate that the real destination may not be located beyond itself. To address this problem, we propose that the end node (i.e., the proxy) in the RR-ZAP route has to occasionally send back dummy packets that have the same size as the ACK/NACK to confuse the attacker.

The disclosed position of a node makes it prone to a so-called *target-oriented attack*, under which an attacker can stay close to its target node and monitor its behavior. The target may probably be the destination if the attacker finds that after the target sends a short message, a private route is generated and the target appears in the AZ. To mitigate such an attack, background noise is needed. A node can occasionally send out dummy packets that have the same pattern as requests and ACKs. In this case, when a real request or ACK is sent, the attacker cannot be certain. The injection of background noise also is the only solution for mitigating a global attack, which is less likely in real networks.

Sybil attacks are attacks under which a node transmits packets using different identities. Since our scheme does not configure its parameters based on the number of heard nodes, such an attack has no impact on the anonymity achieved by ZAP.

4.2 Intersection Attack: The Impact of Node Mobility on Anonymity

An *intersection attack* occurs when an attacker knows that its Entity of Interest (EOI) is in more than one anonymity set. In this case, it concludes that the EOI must be in the intersection of all those anonymity sets. As the intersection set is smaller than any of the original set, the anonymity level of the EOI decreases.

In ZAP, node mobility helps attackers carry on intersection attacks and therefore degrade node anonymity. This is especially the case when the communication between the source and the destination lasts for a long time. Fig. 5 shows an example of such an attack. Suppose that two packets arrive at the AZ at times t_1 and t_2, respectively. As the packets have the same AZ, the attacker knows that they are for the same destination. At time t_1, a set_1 of nodes is located in the AZ, and at time t_2, a set_2 of nodes is located in the AZ. The sets set_1 and set_2 are not equal, because some nodes may have moved out or into the AZ between the two transmissions. To an attacker, the anonymity set for the destination includes only the nodes that are in the AZ at both times t_1 and t_2, that is, the intersection of the anonymity sets at t_1 and t_2. In this example, it is easy for the attacker to infer that the destination node is either e or f. The size of the anonymity set is reduced to 2, instead of 6, for set_1 or 5 for set_2.

If a session lasts long, the number of nodes remaining in the AZ can become small. The destination anonymity can thus become very low. Note that the nodes, which were originally out of an AZ and then moved in the AZ during the communication, do not contribute to anonymity, because the attacker knows that these nodes cannot be the destination anyway.

4.3 Mitigating Techniques against Intersection Attack

Different approaches can be adopted to mitigate the impact of node mobility and to reduce the anonymity degradation. One approach is to break a long-duration session into a number of short subsessions that use different AZs. For each subsession, a new PD and the corresponding symmetric key are generated. As a subsession does not last a long time, the destination anonymity may only decrease moderately because of mobility. The challenge is how these subsessions can be made unlinkable. A straightforward solution is to increase and randomize the intersubsession durations, which improves anonymity at the cost of communication delay.

Another approach is to expand the AZ as time elapses. Such an approach can be applied to G-ZAP or RR-ZAP by expanding the geocasting region or adding more redundant hops. Taking a G-ZAP session as an example, the source can increase the size of the D-AZ to maintain a certain anonymity level. A node that is moving away from the original D-AZ may still be in the expanded D-AZ. The impact of node mobility on anonymity degradation is compensated. When or how fast the D-AZ should be expanded can be determined based on the anonymity requirement, node density, and mobility. Note that in this approach, the source does not have to wait for a new connection request from the destination to expand the D-AZ. The cost is the increased communication overhead, because a data packet has to be flooded in a larger D-AZ. Similar operations can be used in RR-ZAP as well.

5 Analysis

In this section, we present an analysis of various design issues that relate to the different ZAP approaches. We build mathematical models for destination anonymity evaluation. We then calculate the packet delivery ratio in PD-ZAP, assessing the impact from the distance parameter, i.e., the distance between the real and the PDs. Finally, we address the problem of how the best destination anonymity in RR-ZAP can be achieved based on the redundant route generation through random matching. When applicable, the analysis assumes a uniform distribution of node density and a random mobility following the model used by many early work.

5.1 Destination Anonymity

In all ZAP approaches, the destination anonymity is determined by the number of nodes that stay in the AZ. Such a value generally decreases as time elapses, because nodes may move out of the AZ. Let $\alpha_{pd}(t)$, $\alpha_g(t)$, and $\alpha_{RR}(t)$ denote these numbers for PD-ZAP, G-ZAP, and RR-ZAP after a communication that has lasted for a time t. They are the destination anonymity measured through the sizes of the resulting anonymity sets. In general, letting the AZ have an area of A and node density be ρ, the destination anonymity is $A \times \rho \times p_{stay}(t)$, where $p_{stay}(t)$ is the probability that a node stays within the area A after time t. $\alpha_{pd}(t)$, $\alpha_g(t)$, and $\alpha_{RR}(t)$ are calculated as follows:

Based on mobility statistical results in [23], the probability $p_{stay}(t)$ for a randomly moving node[2] with an average speed of $E[v]$ to stay in a region that has an area of A and a perimeter of L after a time t is exponentially distributed, and

$$p_{stay}(t) = e^{-t/\bar{t}}. \tag{1}$$

Here, $\bar{t}$ is calculated by

$$\bar{t} = \frac{\pi \times A}{L \times E[v]}. \tag{2}$$

In particular, when the region is a circular area with a radius of r,

$$\bar{t} = \frac{\pi r}{2E[v]}. \tag{3}$$

Let d_τ and l_τ be the distance threshold values used in PD-ZAP and RR-ZAP, respectively, and R_{AZ} be the radius for the D-AZ in G-ZAP. For $\alpha_{pd}(t)$ and $\alpha_g(t)$, using (1), (2), and (3), we have

$$\alpha_{pd}(t) = \rho\pi d_\tau^2 e^{\frac{-2tE[v]}{\pi d_\tau}}, \tag{4}$$

$$\alpha_g(t) = \rho\pi R_{AZ}^2 e^{\frac{-2tE[v]}{\pi R_{AZ}}}. \tag{5}$$

Let R be the ad hoc radio transmission range. Considering RR-ZAP, let S_{RR} and L_{RR} be the area and the perimeter for the AZ and let d be the distance between the source and the destination. Then

$$\alpha_{RR}(t) = \rho S_{RR} e^{\frac{-tE[v]L_{RR}}{\pi S_{RR}}}, \tag{6}$$

where

$$\begin{aligned} S_{RR} = {} & \pi\left(d_\tau^2 + R^2\right) + 2l_\tau d - l_\tau\left(\sqrt{d_\tau^2 - l_\tau^2} + \sqrt{R^2 - l_\tau^2}\right) \\ & - \left(d_\tau^2 \arcsin\frac{l_\tau}{d_\tau} + R^2 \arcsin\frac{l_\tau}{R}\right), \end{aligned} \tag{7}$$

$$\begin{aligned} L_{RR} = {} & 2\pi(d_\tau + R) + 2d - 2\left(d_\tau \arcsin\frac{l_\tau}{d_\tau} + R\arcsin\frac{l_\tau}{R}\right) \\ & - 2\left(\sqrt{d_\tau^2 - l_\tau^2} + \sqrt{R^2 - l_\tau^2}\right). \end{aligned} \tag{8}$$

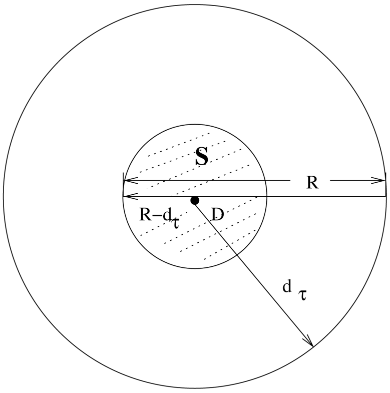


Fig. 6. Analysis on a packer delivery failure in PD-ZAP.

5.2 Impact of Distance Threshold Value on Data Delivery in PD-ZAP

We estimate the lower bound of the probability that the destination receives the data packet when d_τ has different values. We assume that packets are forwarded toward the PD at D. The D-AZ then is the circular area that is centered at D and has a radius of d_τ. If the node closest to D, which will later be the proxy for packet delivery, is not farther than $R - d_\tau$ from D, then the destination located in the D-AZ will certainly receive the packets. The probability that the destination will receive the packet is then the same as the probability that there is at least one node in the shaded area $\mathcal{S}$, as shown in Fig. 6, which is centered at D and has a radius of $R - d_\tau$.

It is well known that if the node density is equal to ρ, the probability that there is at least one node different from the destination in the area $\mathcal{S}$, denoted as $p_{n\geq 1}$, is expressed as

$$p_{n\geq 1} = 1 - e^{-\rho S} = 1 - e^{-\rho\pi(R-d_\tau)^2}. \tag{9}$$

Because nodes are mobile, the destination that can originally receive a message from the proxy may not be able to receive it after they have moved away from each other. The probability that after a time t, the destination still stays in the D-AZ can be obtained through (1), (2), and (3). In summary, at different d_τ, the probability that after a time t, the destination is still able to receive the packet from a proxy, denoted as $p_{succ}(d_\tau, t)$, can be formulated as

$$p_{succ}(d_\tau, t) = \left(1 - e^{-\rho\pi(R-d_\tau)^2}\right) \times e^{\frac{-2tE[v]}{\pi d_\tau}}. \tag{10}$$

This is the lower bound, because a destination may receive a packet, even if 1) the proxy is not located in $\mathcal{S}$ and 2) the destination has moved out of the D-AZ. The specific case depends on where the destination is located and from which direction the packet is forwarded toward the PD.

5.3 Redundant Route Generation in RR-ZAP

In order to simplify the presentation, we use hop counts to denote the length of a path, and we address the issue of the PD selection as the problem of extending an N-hop PD-ZAP route to an N'-hop RR-ZAP route, where $N' \geq N$. In the example in Fig. 4, $N = 4$, and $N' = 6$. We denote the maximum hop count for any route in the network as N_{max}. We examine the random match and determine for an

[2] The random-moving model is the moving pattern under which a mobile user may change direction and speed at any time. Random walk can be looked as a particular pattern that belongs to random moving.

N-hop route, which is the probability that it should be extended to an N'-hop redundant route, where N and N' are different numbers not greater than N_{max}.

The N'-hop redundant route results in the best anonymity for the destination if, based on network information, the attacker can only draw the conclusion that the probabilities that N is equal to $1, 2, \cdots, N'$ are all the same, as this leads to the highest entropy that can be used for evaluating the anonymity [24].

Assume that in the network, the probability that a PD-ZAP route has a hop count of N is $P\{N\}$, and this information is known to all the nodes in the network, including the attackers. Let $P\{N \rightarrow N'\}$ denote the probability that when a PD-ZAP route has a hop count of N, after the random match, the hop count for the RR-ZAP route is N'. Let $P\{N|N'\}$ denote the probability that when a N'-hop RR-ZAP route appears, the hop count for the PD-ZAP route is N. Based on the above analysis, the RR-ZAP route should have the following desired property:

$$\begin{aligned} P\{1|N'\} &= P\{2|N'\} \\ &= \cdots P\{L|N'\} \cdots P\{K|N'\} \cdots = P\{N'|N'\}, \end{aligned} \tag{11}$$

where $L < K \leq N'$.

According to the Bayesian Theorem,

$$P\{L|N'\} = \frac{P\{L\}P\{L \rightarrow N'\}}{P\{N'\}}, \tag{12}$$

$$P\{K|N'\} = \frac{P\{K\}P\{K \rightarrow N'\}}{P\{N'\}}. \tag{13}$$

To make $P\{L|N'\} = P\{K|N'\}$, the following condition must be satisfied:

$$\frac{P\{L \rightarrow N'\}}{P\{K \rightarrow N'\}} = \frac{P\{K\}}{P\{L\}}, \tag{14}$$

along with the conditions

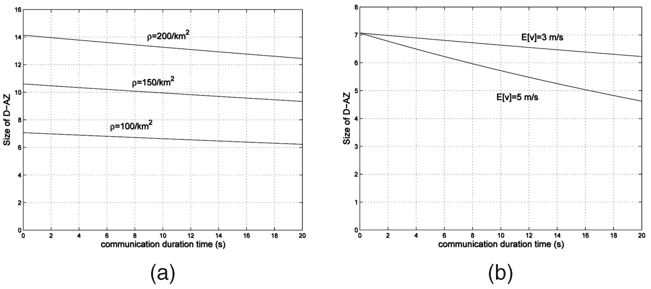
$$\sum_{N'=max(N,N_{min})}^{N_{max}} P\{N \rightarrow N'\} = 1, \tag{15}$$

$$\sum_{N=1}^{N_{max}} P\{N\} = 1. \tag{16}$$

Note that the above random-matching rule applies only when the probability that a PD-ZAP route has a maximum hop count is small. For example, when $K = N' = N_{max}$, $P\{K \rightarrow N'\} = 1$. If $P\{K\}$ (which is equal to $P\{N_{max}\}$) is large, it is difficult to satisfy (14). In this case, (14) has to be adjusted, and the rule for maximizing the entropy value is applied, which can be expressed as

$$max\left(-\sum_{i=N_{min}}^{N_{max}} \sum_{j=1}^{i} P\{j|i\} \log P\{j|i\} \right). \tag{17}$$

To obtain the maximum value from (17), an exhausting search can be used, of which the computing load depends on the grid of probability searching step.



(a) (b)

Fig. 7. Destination anonymity versus node density and mobility. (a) Node density. (b) Mobility.

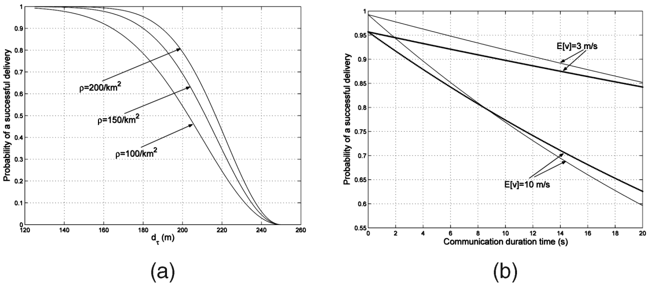
6 Performance Evaluation

In this section, we present both numerical results and simulation evaluations. The numerical results are obtained based on the analysis presented in Section 5, which illustrate the trends of changes. The simulation performs detailed packet-level experiments to further evaluate the three variants of ZAP.

6.1 Numerical Results

The numerical results use an average speed of 1.5 m/s and a node density of $100\ \mathrm{nodes/km}^2$, unless otherwise specified. Fig. 7 shows the destination anonymity for destination-based ZAP approach, i.e., the number of nodes in the D-AZ. We consider PD-ZAP, with the distance threshold value d_τ being 150 m. Fig. 7a shows the size of the AZ at different densities. PD-ZAP can then be used if the anonymity requirement is not high. Fig. 7b shows that the anonymity decreases faster if the speed increases, because a node that is originally in the D-AZ may move out of it in a shorter time. This means that when nodes have high mobility, a source may need to start a new session more frequently to maintain the required anonymity level. The results for G-ZAP and RR-ZAP reflect similar trends. The evaluation methodology is to set the D-AZ to different shapes.

Fig. 8 reports a lower bound for the probability of a successful packet delivery in the PD-ZAP protocol. As shown in Fig. 8a, such a probability is relatively high when d_τ is not very large (i.e., $d_\tau \leq 150\ \mathrm{m}$). The probability also increases as the node density increases. Fig. 8b shows the impact of mobility. The thin lines represent the results for $d_\tau = 125\ \mathrm{m}$, while the bold lines represent the results for $d_\tau = 150\ \mathrm{m}$. We can observe that when node speed increases, the probability of a successful delivery decreases



(a) (b)

Fig. 8. Lower bound for the probability of a successful delivery in PD-ZAP. (a) Without mobility. (b) Impact of mobility.

faster as time elapses, because the destination may move out of the D-AZ in a shorter time.

Based on the numerical results, it is observed that in networks where the node density is not low, a decent anonymity can be preserved by the simplest PD-ZAP protocol. Under a reasonable node mobility, this anonymity will not degrade significantly in a short time.

6.2 Simulation Study

We further evaluate the destination anonymity and the network performance of the proposed protocols through simulation. The evaluation metrics include the following:

1. *Size of the anonymity set.* The number of nodes that remain in the anonymous zone when a session ends compared to those at the beginning of the session (note that this metric evaluates the intersection attack).
2. *Packet delivery ratio.* The ratio between the number of data packets received and those originated by the sources.
3. *Normalized packet forwarding overhead.* The number of packets transmitted by ZAPs normalized to those transmitted by GPSR under the same condition.
4. *Average end-to-end packet latency.* The average time from when the source generates the data packet to when the destination receives it.

We evaluate protocols PD-ZAP, G-ZAP, G-ZAP with region-based efficient flooding (G-ZAP-RBF; refer to Section 3.2), and RR-ZAP. For RR-ZAP, the simulation area limits the number of hops that we can choose for redundancy. Thus, in our implementation, a PD is positioned at the boundary of the simulated field. The location is randomly selected from a segment centered at the intersection of the boundary and the extended link from the source to the destination. The length of the segment is chosen to ensure that l_τ is equal to half the transmission range. We present GPSR for reference when appropriate.

We use QualNet [25], a detailed packet-level network simulator, to investigate the impact of the protocol specific parameters and varying network conditions on the aforementioned metrics. The simulated ad hoc network has 180 nodes with uniform initial distribution. The servers are part of the network and participate in communications as sources. The nodes move according to a Random Waypoint Model [32], with the pause time being zero and the minimum and the maximum speeds being set to the same value (note that this configuration avoids the speed decaying problem [26]). The average density is around 20 neighbors per node. Simulations use renewal CBR application in order to constantly maintain five CBR sessions. Each source generates data packets of 256 bytes at a rate of 4 packets per second. The source-destination pairs are chosen randomly from all the nodes (but we exclude the pairs that have the destination located close to the edge of the network in order to avoid artificial degradation in anonymity). We use the IEEE 802.11b DCF at the MAC layer with a link bandwidth of 2 megabits per second (Mbps). We use the default radio power range of 370 m (according to the default parameters for radio and propagation models in QualNet). In consideration of the difference from a real wireless scenario, we match the protocol parameters d_τ, l_τ, and R_{AZ} to this value. For example, d_τ and R_{AZ} are set to 370 m. The field size is also matched to be 2,000 m × 2,000 m. Detailed case-specific parameter values are given with the figures. The results are averaged over several simulation runs with various random seeds.

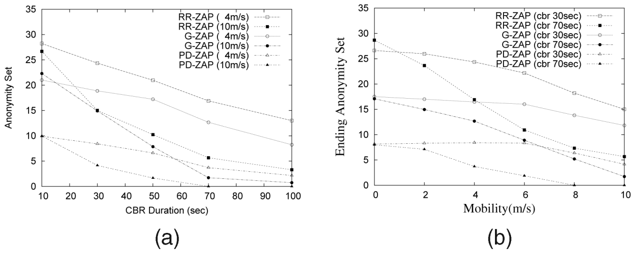


Fig. 9. Anonymity set. (a) AS versus CBR duration. (b) AS versus mobility.

6.2.1 Anonymity

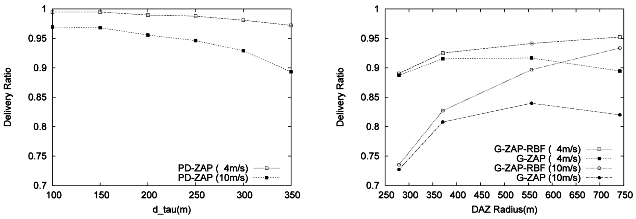
The destination anonymity is measured by the size of the anonymity set $Size_{AS}$, which consists of the nodes remaining in the AZ throughout the session. We investigate how it is affected by the CBR session time, mobility, and anonymous zone configuration. The default AZ sizes are 250 m for d_τ in PD-ZAP and 370 m for R_{AZ} in G-ZAP.

Fig. 9a reports the change in $Size_{AS}$ as a function of the session duration. The experimental results reported in the figure show several interesting facts. First, in general, the anonymity set of RR-ZAP is larger than that of PD-ZAP and G-ZAP, because the entire route becomes the anonymous region, which, in most cases, is larger than a destination-based D-AZ. Moreover, given that R_{AZ} is larger than d_τ, the AS size of G-ZAP is larger than that of PD-ZAP. Second, when the session duration increases, all curves show a decrease in the size of the anonymity set. Third, when mobility is high, the anonymity set size decreases faster, because more nodes move out of the initial anonymous zone during the session. Note that under similar node density and mobility, the anonymity for PD-ZAP is close to what is obtained through analysis (refer to Fig. 7a).

Fig. 9b shows the change in $Size_{AS}$ of ZAPs as a function of mobility for long and short sessions. The trends are similar to the ones shown in the previous figure. RR-ZAP has the largest AS size. But, when mobility increases, the size decreases more quickly than those of the other two protocols, especially when sessions last longer. The reason is that RR-ZAP's anonymous zone is generally long and narrow, and it is thus more sensitive to mobility. Both G-ZAP and PD-ZAP can tolerate higher mobility when the session is short (30 seconds). Up to mobility values equal to 6 m/s, the sizes of the AS are mostly not affected by mobility due to the fact that few nodes can move out of the original AS region in a short period of time. When the session is long (70 seconds), all the ZAPs start degrading at low mobility values.

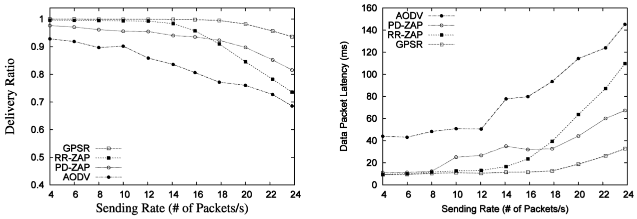
6.2.2 Routing Performance

We investigate how the packet delivery performance of the ZAP protocols is affected by session time, mobility, sizes of



(a) (b)

Fig. 10. D-AZ's impact on delivery ratio. (a) PD-ZAP. (b) G-ZAP.



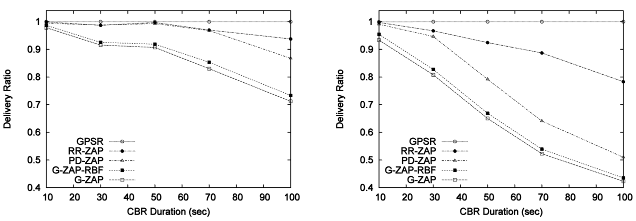
(a) (b)

Fig. 12. Impact of traffic load. (a) Delivery ratio. (b) Latency.

the D-AZs, and traffic load. While we try to stress one condition, we keep other parameters moderate.

Fig. 10 investigates how the AZ configurations affect PD-ZAP, G-ZAP, and G-ZAP-RBF with respect to the delivery ratio, respectively. Sessions are kept short at 30 seconds. Fig. 10a shows that PD-ZAP maintains high delivery ratio when mobility is low (4 m/s) no matter how d_τ increases. This is because the distance that a node can move in the short session time does not cause many nodes to move out of its D-AZ, which is a little smaller than a node's transmission range. However, the delivery ratio degrades quickly in the case of high mobility (10 m/s), as expected. Fig. 10b first illustrates that G-ZAP-RBF can achieve better performance than G-ZAP because of the efficient flooding in the AZ. When the zone size increases, the advantage of using region-based flooding becomes more evident. Notice that in our other results, we configured R_{AZ} to be 370 m for G-ZAP. In such a case, G-ZAP and G-ZAP-RBF have the same behavior, as shown in the figure. Fig. 10b also shows that a small R_{AZ} can lower the packet delivery ratio, as a node can easily move out of the zone. On the other hand, a too large R_{AZ} could also reduce packet delivery ratio. The reason is that a larger zone incurs more data broadcast and longer hops to deliver the packets to the real destination. Given that packets are broadcast without RTS/CTS protection from hidden terminals, a higher number of collisions occur. The figure also shows that higher mobility results in a lower delivery ratio.

Fig. 11 reports the impact of session duration, when d_τ is 250 m for PD-ZAP and the R_{AZ} is 370 m for G-ZAP. The figure shows that GPSR has a nearly perfect data delivery ratio over all the session lengths (location updates in GPSR simulation are instantaneous). PD-ZAP and RR-ZAP perform very close to GPSR when sessions are not very long. They suffer from delivery ratio degradation when sessions are long. High mobility has a large impact, even if sessions are short. The impact of session duration and mobility results from the fact that destination nodes move away from the anonymous region. RR-ZAP greatly improves performance compared to PD-ZAP in high mobility situations due to the following reason. In PD-ZAP, a real destination only listens to one proxy, while in RR-ZAP, the real destination has the chance to overhear from more than one node en route. Even when a node moves away from a previous geo-forwarding node, it could move closer to another geo-forwarding node. The figure shows that this advantage is significant.



(a) (b)

Fig. 11. Session duration impact on delivery ratio. (a) Low mobility. (b) High mobility.

Fig. 11 also suggests that G-ZAPs perform worse than PD-ZAP. The reason for this is when all the packets use geo-forwarding, the center area of the simulation field is more heavily loaded than the edge area. While a normal ad hoc routing protocol may find a way of detouring, geo-forwarding always tries to send packets along the straight line from the source to the destination. When G-ZAP floods data packets within a zone (a hot spot) that covers center the area, it causes congestion. A hot spot blocks the geo-forwarding path and incurs more packet collisions if overlapping with other hop spots. In addition, data flooding has no reliability mechanism as well. All these factors account for a lower packet delivery ratio of G-ZAP than that of PD-ZAP. In this figure, the G-ZAP-RBF has limited advantage over G-ZAP, because R_{AZ} is small.

We also study the impact of traffic load on the routing performance. In the simulation, node mobility is 2 m/s. There are five concurrent CBR sessions at any time, and the session duration is 10 seconds. The load is increased by increasing the CBR sending rate (from 8 to 48 kilobits per second (Kbps)). We show simulation results of the protocols GPSR, RR-ZAP, PD-ZAP, and AODV. For PD-ZAP, d_τ is 250 m. Other parameters are set to the default values mentioned before.

Fig. 12a gives the trend of data delivery ratio when the data sending rate increases. We have the following observations. First, as expected, all protocols experience degradation when the sending rate increases due to increased contention and congestion during wireless transmission. GPSR has the best delivery ratio because of its lightweight geographical routing and the fact that it is always able to find the next hop (given the simulated nodal density). AODV results in the worst performance, as it requires route discovery (which involves route request flooding and route reply unicast) before data transmission, and the stability of established routes is prone to node mobility over time. Second, consistent with the previous results, when the sending rate is low, RR-ZAP performs like GPSR and is able to deliver more data packets than PD-ZAP. The advantage of RR-ZAP is derived from the fact that the destination has more chances to receive data from adjacent nodes along the path between the source and the PD. In contrast to PD-ZAP, a node may only have one opportunity to hear each data packet from the proxy. Third, when the sending rate increases, RR-ZAP shows a faster degradation trend than the others. The reason is that many extra data are forwarded along the redundant routes. In addition, in the current simulation scenario, redundant routes are extended to the field boundaries, which creates more cases of cross traffic flows. As a result of the interference among the traffic flows, the probability that a destination can overhear a transmission decreases.

Fig. 12b shows the impact of data sending rate on the end-to-end data delivery latency. All protocols show an increasing long latency due to increased channel contentions when the load increases. AODV has the longest latency because of the initial route discovery process, and GPSR has the smallest latency due to the lightweight geographical forwarding. PD-ZAP has a longer latency than GPSR, because the overall path that a packet takes could be longer than the geographical shortest path due to the last-hop anonymous transmission. For RR-ZAP, when the sending rate becomes higher, the impact of the increased congestion and contention on RR-ZAP tends to be stronger than in the other protocols.

In summary, our simulations show that for destination anonymity protection, both RR-ZAP and G-ZAP are able to successfully increase the AS size. RR-ZAP, however, is more sensitive to mobility and communication duration than other ZAPs, while G-ZAP can tolerate higher mobility than other ZAPs. For routing performance, RR-ZAP has close performance (high packet delivery and low packet latency) to GPSR in most cases due to the increased opportunities in overhearing transmissions. The results further demonstrate the trade-offs between the anonymity set size and the performance. The RR-ZAP has the best balance in terms of both protection and routing performance. On the other hand, G-ZAP can produce stronger protection at the cost of higher overhead and lower packet delivery ratio as compared to PD-ZAP.

7 Related Work

Anonymous communication protocols studied for wired Internet have followed MIX [27] techniques, Onion routing [28], broadcast [29], or multicast [30] approaches. However, they are not always applicable to MANETs. For example, protocols based on MIX [27] techniques and Onion routing [28] require a priori underlying security associations among entities via a fixed infrastructure, which is very difficult in dynamic and self-organizing MANETs. Approaches using broadcast [29] or multicast [30] have high bandwidth demand. The obstacles against achieving communication-end privacy, especially destination anonymity, also come from the fact that in on-demand routing protocols such as AODV [31] and DSR [32], global flooding is required at the route discovery stage. The destination identity is carried in the request; therefore, it is revealed to the entire network. All nodes in the network may thus become aware about communications being established.

To date, several approaches have been reported addressing anonymity in MANETs, including protocols using single routing path [11], [5], [33], [37]. In particular, under the AO2P protocol [11], the destination position is the only position information disclosed in the network for routing. As in traditional positioning algorithms, in AO2P, the route is discovered by delivering a routing request from the source toward the position of the destination. However, AO2P does not rely on the local position information exchange. The other relevant approach is the untraceable on-demand routing protocol called ANODR [5]. This protocol uses an onion structure for routing discovery. To reduce the cost and latency of the encryption/decryption, a symmetric key based on *Boomerang Onions* is used. Once a route is discovered, pseudorandom numbers are used as temporary IDs for each link along the route. Each node only knows the pseudo numbers for its previous hop and next hop. Communication privacy is achieved, because real IDs are not revealed.

In the MASK protocol [33], a neighborhood authentication protocol that allows neighboring nodes to authenticate each other without revealing their identities is designed for communication anonymity. However, the approach needs to reveal the destination ID for on-demand route discovery. Therefore, only a conditional anonymity can be achieved for the destination; that is, a tracer knows which node is the destination, yet the tracer does not know where the destination is. Source anonymity for ad hoc routing has been investigated by Yang et al. [37]. Under their approach, each intermediate node substitutes the source identity in a route request with its own identity. Layered encryption is used in route reply packets to set up onions for data communications. The protocol does not hide each other's ID when sending/receiving routing messages nor the destination's identity in the flooding. Thus, within a neighborhood, no protection is ensured for identity anonymity.

Non-single-path approaches are introduced in [38] and [39]. In [38], a packet coding technique is used to combine multicast and onion routing to address the anonymity threats coming from both the global and local adversaries. The source and destination anonymity is protected through forming multicast forests, and unlinkability is ensured by onion-based packet encryption. The need for global routing information helps build a multicast forest but limits its use to high-bandwidth networks. Multipath routing is used in [39], where data traffic is split to and forwarded through multiple paths randomly over time. To evaluate the traffic privacy, an entropy metric is designed to characterize the limited knowledge that a single intermediate node can gather. The scheme employs source routing where the source will build a path poll for traffic splitting. Thus, it is designed for static mesh network with a fixed infrastructure.

Geographic zone has been used for protecting user privacy. For example, *Mix Zone* [35] is designed to prevent application servers from tracing their users' movement. Within a mix zone, a mobile user obtains LBSs without revealing its accurate position. In [40], an anonymity proxy enables spatial and temporal anonymity by perturbing the location data in terms of transmitting time and position information, enabling location privacy in using LBSs. While the above two work are middleware solutions, the *Motion Mix* approach [36] explores mobility for protection. It uses the geographic area generated due to the attacker's inability to pinpoint a transmission. Any nodes moving in and out of the motion mix becomes a protection against nearby eavesdroppers. Geographic zone in ZAP is used differently from these approaches because of the different attacker model and design goal.

8 Conclusion

In this paper, we have proposed ZAP, an anonymous georouting protocol that adopts fuzzy positions to create AZ for destination anonymity. Nodes residing in the AZ form the anonymity set, which protects the real destination. Because nodes are mobile, the anonymity set in our protocols is dynamic, unlike the case of wired networks. We have introduced a set of strategies that effectively increase the anonymity set: the PD-ZAP, G-ZAP, and RR-ZAP protocols. We use both analysis and simulation to study various performance aspects such as node anonymity and packet delivery percentage. We have found that RR-ZAP, which uses redundant routes to improve anonymity, can achieve a high packet delivery ratio and assure the highest anonymity. If the anonymity requirement is not high, PD-ZAP can be used, because it achieves efficient node anonymity and still achieves a good routing performance in many cases. G-ZAP has to trade the performance for anonymity. The main problem arises from contentions in the AZ. Thus, it is suitable for low-density networks, for which the other approaches are less effective.

References

[1] G. Resta, P. Santi, and J. Simon, "Analysis of Multi-Hop Emergency Message Propagation in Vehicular Ad Hoc Networks," *Proc. ACM MobiHoc,* 2007.

[2] Y.-C. Hu and H.J. Wang, "A Framework for Location Privacy in Wireless Networks," *Proc. ACM SIGCOMM Asia Workshop,* 2005.

[3] Q. He, D. Wu, and P. Khosla, "Quest for Personal Control over Mobile Location Privacy," *IEEE Comm. Magazine,* vol. 42, no. 5, pp. 130-136, 2004.

[4] P. Kamat, Y. Zhang, W. Trappe, and C. Ozturk, "Enhancing Source-Location Privacy in Sensor Network Routing," *Proc. 25th IEEE Int'l Conf. Distributed Computing Systems (ICDCS),* 2005.

[5] J. Kong and X. Hong, "ANODR: Anonymous on Demand Routing with Untraceable Routes for Mobile Ad-hoc Networks," *Proc. ACM MobiHoc '03,* June 2003.

[6] D. Mohapatra and S.B. Suma, "Survey of Location-Based Wireless Services," *Proc. IEEE Int'l Conf. Personal Wireless Comm. (ICPWC),* 2005.

[7] X. Wu and H. Schulzrinne, "Location-Based Services in Internet Telephony," *Proc. Second IEEE Consumer Comm. and Networking Conf. (CCNC '05),* Jan. 2005.

[8] I. Stojmenovic, "Position-Based Routing in Ad Hoc Networks," *IEEE Comm. Magazine,* vol. 40, no. 7, pp. 128-134, July 2002.

[9] T. Camp, J. Boleng, B. Williams, L. Wilcox, and W. Navidi, "Performance Comparison of Two Location-Based Routing Protocols for Ad Hoc Networks," *Proc. IEEE INFOCOM '02,* pp. 1678-1687, 2002.

[10] B. Karp and H.T. Kung, "GPSR: Greedy Perimeters Stateless Routing for Wireless Network," *Proc. ACM MobiCom,* 2000.

[11] X. Wu and B. Bhargava, "AO2P: Ad Hoc On-Demand Position-Based Private Routing," *IEEE Trans. Mobile Computing,* 2005.

[12] R.P. Minch, "Privacy Issues in Location-Aware Mobile Devices," *Proc. 37th Ann. Hawaii Int'l Conf. System Sciences (HICSS),* 2004.

[13] B. Gedik and L. Liu, "Location Privacy in Mobile Systems: A Personalized Anonymization Model," *Proc. 25th IEEE Int'l Conf. Distributed Computing Systems (ICDCS),* 2005.

[14] R. Cheng, D.V. Kalashnikov, and S. Prabhakar, "Querying Imprecise Data in Moving Object Environments," *IEEE Trans. Knowledge and Data Eng.,* vol. 16, no. 9, pp. 1112-1127, Sept. 2004.

[15] M.K. Reiter and A.D. Rubin, "Crowds: Anonymity for Web Transactions," *ACM Trans. Information and System Security,* vol. 1, no. 1, pp. 6-92, 1998.

[16] *The Keyed-Hash Message Authentication Code,* FIPS 198, Nat'l Inst. Standards and Technology (NIST), 2002.

[17] K. Chen and K. Nahrstedt, "Effective Location-Guided Tree Construction Algorithms for Small Group Multicast in MANET," *Proc. IEEE INFOCOM,* 2002.

[18] R. Ganhdi, S. Parthasarathy, and A. Mishra, "Minimizing Broadcast Latency and Redundancy in Ad Hoc Networks," *Proc. ACM MobiHoc,* 2003.

[19] W. Luo and J. Wu, "On Reducing Broadcast Redundancy in Ad Hoc Wireless Networks," *IEEE Trans. Mobile Computing,* vol. 1, no. 2, pp. 111-122, 2002.

[20] Y. Tseng, S. Ni, Y. Chen, and J. Sheu, "The Broadcast Storm Problem in a Mobile Ad Hoc Network," *Proc. ACM MobiCom '99,* Aug. 1999.

[21] X. Wu, "VPDS: Virtual Home Region based Distributed Position Service in Mobile Ad Hoc Networks," *Proc. 25th IEEE Int'l Conf. Distributed Computing Systems (ICDCS),* 2005.

[22] A. Pfitzmann and M. Kohntopp, "Anonymity, Unobservability, and Pseudonymity—A Proposal for Terminology," *LNCS 2009,* H. Federrath, ed., p. 19, 2000.

[23] R. Thomas, H. Gilbert, and G. Mazziotto, "Influence of the Moving of the Mobile Stations on the Performance of a Radio Cellular Network," *Proc. Third Nordic Seminar,* 1988.

[24] A. Serjantov and G. Danezis, "Towards an Information Theoretic Metric for Anonymity," *Proc. Second Int'l Workshop Privacy Enhancing Technologies (PET),* 2002.

[25] *QualNet.* Scalable Network Technologies (SNT), http://www.qualnet.com/, 2008.

[26] J. Yoon, M. Liu, and B. Noble, "Random Waypoint Considered Harmful," *Proc. IEEE INFOCOM,* 2003.

[27] D.L. Chaum, "Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms," *Comm. ACM,* vol. 24, no. 2, pp. 84-88, 1981.

[28] M. Reed, P. Syverson, and D. Goldschlag, "Anonymous Connections and Onion Routing," *IEEE J. Selected Areas in Comm.,* special issue on copyright and privacy protection, vol. 16, no. 4, pp. 482-494, 1998.

[29] R. Sherwood, B. Bhattacharjee, and A. Srinivasan, "P5: A Protocol for Scalable Anonymous Communication," *Proc. IEEE Symp. Security and Privacy (S&P '02),* pp. 53-65, May 2002.

[30] V. Scarlata, B. Levine, and C. Shields, "Responder Anonymity and Anonymous Peer-to-Peer File Sharing," *Proc. Ninth IEEE Int'l Conf. Network Protocols (ICNP),* 2001.

[31] C.E. Perkins and E.M. Royer, "Ad Hoc On-Demand Distance Vector Routing," *Proc. Second IEEE Workshop Mobile Computing Systems and Applications (WMCSA '99),* pp. 90-100, 1999.

[32] D. Johnson and D. Maltz, "Dynamic Source Routing in Ad Hoc Wireless Networks," *Proc. ACM SIGCOMM—Computer Comm. Rev.,* 1996.

[33] Y. Zhang, W. Liu, and W. Luo, "Anonymous Communications in Mobile Ad Hoc Networks," *Proc. IEEE INFOCOM,* 2005.

[34] X. Wu, J. Liu, X. Hong, and E. Bertino, "Achieving Anonymity in Mobile Ad Hoc Networks Using Fuzzy Position Information," *Proc. Second Int'l Conf. Mobile Ad Hoc and Sensor Networks (MSN '06),* Dec. 2006.

[35] A.R. Beresford and F. Stajano, "Mix Zones: User Privacy in Location-Aware Services," *Proc. Second IEEE Ann. Conf. Pervasive Computing and Comm. Workshops (PerCom),* 2004.

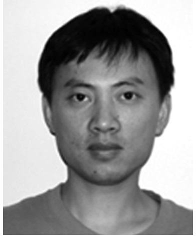
[36] J. Kong, D. Wu, X. Hong, and M. Gerla, "Mobile Traffic Sensor Network versus Motion-MIX: Tracing and Protecting Mobile Wireless Nodes," *Proc. Third ACM Workshop Security of Ad Hoc and Sensor Networks (SASN '05),* pp. 97-106, 2005.

[37] L. Yang, M. Jakobsson, and S. Wetzel, "Discount Anonymous on Demand Routing for Mobile Ad Hoc Networks," *Proc. Second Int'l Conf. Security and Privacy in Comm. Networks (SecureComm),* 2006.

[38] I. Aad, C. Castelluccia, and J.-P. Huubaux, "Packet Coding for Strong Anonymity in Ad Hoc Networks," *Proc. Second Int'l Conf. Security and Privacy in Comm. Networks (SecureComm '06),* pp. 1-10, Aug. 2006.

[39] T. Wu, Y. Xue, and Y. Cui, "Preserving Traffic Privacy in Wireless Mesh Networks," *Proc. Int'l Symp. World of Wireless, Mobile and Multimedia Networks (WoWMoM '06),* June 2006.

[40] M. Gruteser and D. Grunwald, "Anonymous Usage of Location-Based Services through Spatial and Temporal Cloaking," *Proc. ACM MobiSys '03,* May 2003.



Xiaoxin Wu received the BE degree from Beijing University of Posts and Telecommunications in 1990 and the PhD degree from the University of California, Davis, in 2001. Since 2002, he has been a postdoctoral researcher in the Department of Computer Science, Purdue University, working on wireless network privacy and security under a research fellowship from the Institute for Information Infrastructure Protection (I3P). In 2006, he joined Communications Technology Lab, Intel China Research Center, Beijing, working on security and networking issues in WiMax and digital health. His research interests include designing and developing architecture, algorithm, and protocols for network performance improvement in different wireless networks and integrated networks.

Jun Liu received the BE degree in computer science from Beijing (Northern) Jiaotong University in 1998, the MS degree in computer science from Loyola University Chicago in 2001, and the PhD degree in computer science from the University of Alabama in 2007. He is currently a senior analyst in Manhattan Associates. His research interests include protocol design, performance evaluation, security and privacy, routing, and power-aware solutions for wireless mobile networks and sensor networks.

Xiaoyan Hong received the BS and ME degrees in computer science from Zhejiang University, Hangzhou, China, and the PhD degree in computer science from the University of California, Los Angeles, in 2003. She is currently an assistant professor in the Department of Computer Science, University of Alabama. Her research interests include network protocol design, performance evaluation and implementation for multihop, mobile and wireless networks, and wireless sensor networks. Her current research is focused on mobility, privacy, security, routing, and monitoring issues.

Elisa Bertino is currently a professor of computer science and electrical and computer engineering in the Department of Computer Science, Purdue University, where she is also the research director of the Center for Education and Research in Information Assurance and Security (CERIAS). She was a faculty member in the Department of Computer Science and Communication, University of Milan, where she directed the DB&SEC Laboratory. She has been a visiting researcher at the IBM Research Laboratory (now Almaden), San Jose, at the Microelectronics and Computer Technology Corp., at Rutgers University, and at Telcordia Technologies. She is a co-editor in chief of the *Very Large Database Systems (VLDB) Journal*. She also serves on the editorial boards of several scientific journals, including the *IEEE Internet Computing*, *IEEE Security and Privacy*, *ACM Transactions on Information and System Security*, *ACM Transactions on Web*, *Acta Informatica*, *Parallel and Distributed Database Journal*, *Journal of Computer Security*, *Data and Knowledge Engineering*, and *Science of Computer Programming*. She has been a consultant of several companies on data management systems and applications and has given several courses to industries. She has served as a member of the program committee of several international conferences such as the ACM SIGMOD, International Conference on Very Large Data Bases (VLDB), and ACM International Conference on Object-Oriented Programming, Systems, Languages, and Applications (OOPSLA). She was a program cochair of the 14th IEEE International Conference on Data Engineering (ICDE 1998) and the program chair of the 14th European Conference on Object-Oriented Programming (ECOOP 2000), the Seventh ACM Symposium on Access Control Models and Technologies (SACMAT 2002), the Ninth International Conference on Extending Database Technology (EDBT 2004), and the Eighth IEEE Workshop on Policies for Distributed Systems and Networks (POLICY 2007). Her research interests include security, privacy, digital identity management systems, database systems, distributed systems, and multimedia systems. In these areas, she has published more than 250 papers in all major refereed journals and in proceedings of international conferences and symposia. She is a coauthor of *Object-Oriented Database Systems—Concepts and Architectures* (Addison-Wesley, 1993), *Indexing Techniques for Advanced Database Systems* (Kluwer Academic Publishers, 1997), *Intelligent Database Systems* (Addison-Wesley, 2001), and *Security for Web Services and Service-Oriented Architectures* (Springer, Summer 2007). Her research has been sponsored by several organizations and companies, including the US National Science Foundation, the US AirForce Office for Sponsored Research, the I3P Consortium, the European Union (under the Fifth and Sixth IST Research Programmes), IBM, Microsoft, and the Italian Telecom. She is a fellow of the IEEE and the ACM and is a golden core member of the IEEE Computer Society. She received the 2002 IEEE Computer Society Technical Achievement Award for her "outstanding contributions to database systems and database security and advanced data management systems" and the 2005 IEEE Computer Society Tsutomu Kanai Award for her "pioneering and innovative research contributions to secure distributed systems."