A Privacy Preserving Prediction-based Routing Protocol for Mobile Delay Tolerant Networks

Omar Hasan, Jingwei Miao, Sonia Ben Mokhtar, Lionel Brunie
University of Lyon, CNRS, INSA-Lyon, LIRIS, UMR5205, F-69621, France
E-mail: {omar.hasan, jingwei.miao, sonia.benmokhtar, lionel.brunie}@insa-lyon.fr

***Abstract*—A prediction-based routing protocol for mobile delay tolerant networks functions by forwarding a message from one intermediate node to another if the latter has higher probability of encountering the destination node. However, this process compromises the privacy of the nodes by revealing their mobility patterns. In this paper, we propose a privacy preserving prediction-based routing protocol that forwards messages by comparing information about communities of nodes instead of individual nodes. Specifically, it compares the maximum probability that a node in the community of a potential intermediate node will encounter the destination node. We present theoretical security analyses as well as practical performance evaluations. Our simulations on a well established community-based mobility model demonstrate that our protocol has comparable performance to existing prediction-based protocols. Yet our protocol is the only one that preserves the privacy of nodes.**

***Index Terms*—privacy, routing protocols, mobile computing, delay tolerant networks**

I. Introduction

Mobile Delay Tolerant Networks (Mobile DTNs or MDTNs) are wireless mobile networks in which an end-to-end routing path cannot be assumed to exist between the source node and the destination node of a message [1]. A number of networking scenarios have been categorized as Mobile DTNs, such as wildlife tracking sensor networks, Vehicular Ad hoc NETworks (VANETs), Pocket Switched Networks, etc.

In order to deal with the lack of end-to-end connectivity between nodes, message dissemination is often performed in a "store-carry-and-forward" manner [1], where a message is stored by intermediary nodes and forwarded to nodes closer and closer to the destination until it is eventually reached or the message expires. In such scenarios, the mobility pattern of nodes plays an important role in the routing process. Prediction-based routing protocols [2] for Mobile DTNs take advantage of the mobility patterns of nodes to deliver messages. It has been shown that these protocols perform better than other protocols when nodes exhibit well known mobility patterns [3]. For instance, if nodes exhibit social behavior, community-based routing protocols (e.g., Bubble [4]) perform better by reasoning on the membership of the destination node in a community to choose the best relays. Similarly, when nodes have regular movements, regularity-based routing protocols (e.g., Habit [5]) have shown to outperform other protocols by reasoning on the habits of the nodes to choose the best path towards the destination.

Unfortunately, prediction-based routing protocols compromise the privacy of nodes by revealing their mobility patterns. Specifically, they require nodes to exchange and store information about the history of their movements in order to compute the best routing paths. For instance, in Habit [5], nodes exchange their probability of encountering other nodes in the network at given times of the day and days of the week. This allows the source node of a message to select the routing path with the highest probability to reach the destination. Similarly, in PRoPHET [6], nodes exchange their probability to meet the destination node of a message. This allows a node to decide whether it should relay the message through a node that it has encountered. The assumption that nodes will accept to reveal their mobility patterns is not realistic as this information can be used to infer private information about them, as demonstrated by Gambs et al. [7]. Lack of privacy has been identified as one of the reasons for the unwillingness of nodes to participate in Mobile DTNs [8].

In this paper, we present the first Privacy Preserving Prediction-based Routing protocol in Mobile DTNs, named 3PR, which preserves the privacy of node mobility patterns. Specifically, 3PR hides the probability that a node will encounter the destination node of a message. 3PR is intended for environments in which nodes belong to communities. Recent studies of real mobility traces have shown that this is the case for most nodes [4][9].

For routing a message, 3PR distinguishes the routing inside a community from the routing between communities. For disseminating a message inside a community, 3PR relies on the epidemic protocol [10], which by construction preserves the privacy of nodes and is efficient as communities are assumed to be small. The main challenge addressed by 3PR is thus the routing of a message between communities in a privacy preserving manner. To do so, when two nodes from different communities encounter, instead of comparing their respective probabilities to encounter the destination node, they compare the *maximum probability in their community* that a given node will encounter the destination. This probability is periodically computed by nodes belonging to the same community using MDTN-Private-Max and MDTN-Private-Sum protocols. These two protocols, which we propose in this paper, are used for computing the maximum and the sum of a set of values in a privacy preserving manner.

We evaluate 3PR both theoretically by providing security analyses and practically through extensive simulations. We have conducted our simulations based on a well established community-based mobility model [11][12]. We compare the performance evaluation of 3PR against four state-of-the-art protocols, i.e., epidemic [10], Direct [13], PRoPHET [6], and Bubble [4]. Epidemic and Direct are traditionally considered to achieve the upper and lower bounds of routing performance. PRoPHET and Bubble are representatives in prediction-based and social-based routing protocols, respectively. Results show that 3PR has comparable performance to existing prediction-based protocols. Yet, it is the only one that preserves the privacy of nodes.

The remainder of this paper is structured as follows. We first present our system model in Section II. We then present our 3PR protocol in Section III followed by our MDTN-Private-Max and MDTN-Private-Sum protocols presented in sections IV and V respectively. We further present our performance evaluation and related work in sections VI and VII respectively. We conclude the paper in Section VIII.

II. System Model

We consider a set $\mathbb{A}$ of N nodes with communication facilities that can freely roam in a physical environment. The communication facilities consist of a short range wireless connection. Two nodes can communicate only if they are adjacent to each other, i.e, if they are physically within each other's transmission range. We assume that the communication is unreliable, i.e., a message sent from a node to an adjacent node may not arrive. However, we assume that a node knows whether the transmission of a message has been interrupted by a network failure or whether the message correctly reached the intended recipient.

To send a message to a destination node that is not within the transmission range of the source node, the latter uses a routing protocol. The routing strategy that we consider in this work is prediction-based routing [2]. We generalize prediction-based routing protocols as follows: Consider a node a that has a message for a destination node d. When the node a encounters another node b, it forwards a copy of the message to the node b if the probability of b encountering d (given as P_{bd}) is higher than the probability of a encountering d (given as P_{ad}). Thus the probability that a node with a copy of the message will encounter the destination node continues to rise until the message is delivered or the Time To Live (TTL) of the message expires.

As demonstrated in many studies of real human mobility traces, we assume that nodes belong to communities [4]. We define a community C as a set of nodes such that $C \subset \mathbb{A}$. We assume that the nodes in a community are frequently physically collocated and thus a high probability exists of successful message delivery from any source node in a community to any destination node in the community. A node $l \in C$ is designated as the leader of the community. The leader node maintains the list of the nodes in the community. Let the set of nodes in a community $C = \{a_1, a_2, \dots, a_n\}$, where $n = |C|$. We consider a community to comprise of at least three nodes, that is, $n \geq 3$.

We consider the probability that a node a will encounter a node d as private information. Nodes are willing to let this private information be used for routing of messages. However, nodes require that their private information is not revealed to any other node in the network, which includes fellow nodes in a community.

In this paper, we consider the semi-honest adversarial model [14]. The nodes in this model always execute the protocol according to the specification. However, the adversary passively attempts to learn the private information of nodes by using intermediate information gleaned during the execution of the protocols.

III. Privacy Preserving Prediction-based Routing

A. Protocol Description

In this section, we give an overview of 3PR, our Privacy Preserving Prediction-based Routing protocol. A routing example is depicted in Figure 1. This figure shows a number of nodes belonging to three communities C_1, C_2 and C_x. A source node s that belongs to the community C_1 wants to send a message to a node d that belongs to the community C_x.

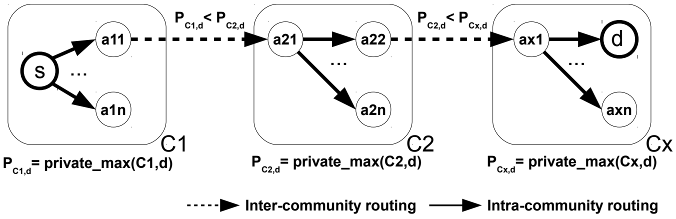


Fig. 1: 3PR Protocol Overview

In 3PR, we distinguish the routing inside a community from the routing between communities. Specifically, when two nodes that belong to the same community encounter each other, they exchange all the messages they have. On the other hand, if two nodes a_{11} and a_{21} that belong to different communities C_1 and C_2 respectively encounter each other, node a_{11} forwards a message intended for a destination node d to node a_{21}, only if a node in community C_2 has a higher probability of encountering d than any given node in community C_1. Let $P_{C_a,d} = max(C_a, d)$ be the maximum probability that a node in community C_a will encounter the destination node d. In Figure 1, when node a_{11} encounters node a_{21}, node a_{11} forwards the message intended for d to node a_{21} because $max(C_2, d) > max(C_1, d)$.

Summarizing, to route a message m from s to d, m is first disseminated in an epidemic manner inside the community C_1. Message m then moves from a community to another such that: (1) at each forwarding step, the probability of a node in the next community to reach the destination is higher than the probability of any given node in the previous community to reach the destination node, (2) as soon as it reaches a community, m is disseminated in an epidemic manner within the community.

A key characteristic of 3PR is that $P_{C_a,d} = max(C_a, d)$, the maximum probability that a given node in community C_a will encounter the destination node d, is computed in a privacy preserving manner, that is without revealing the individual probabilities of the nodes in the community. $max(C_a, d)$ is therefore denoted as $private\_max(C_a, d)$ in Figure 1.

Our protocol 3PR for Privacy Preserving Prediction-based Routing in Mobile DTNs is specified in Figure 2. The computation of $private\_max(C_a, d)$ is performed using a decentralized protocol for privately computing the maximum of a set of values in a delay tolerant manner without revealing the individual values, i.e., MDTN-Private-Max further described in Section IV. The protocol MDTN-Private-Max uses a protocol for privately computing the sum of a set of values in a delay tolerant manner without revealing the individual values, i.e., MDTN-Private-Sum further described in Section V.

The maximum probability is computed periodically in the community independently from the routing protocol. Therefore, the complexity of the MDTN-Private-Max and the MDTN-Private-Sum protocols has no direct impact on the performance of the routing protocol.

Protocol: MDTN-3PR
Participants: Node a and node b, where $a, b \in \mathbb{A}$.
Input: (1) m, a message. (2) d, the destination node of message m. (3) C_a, the set which denotes the community of node a. (4) C_b. (5) $P_{C_a,d} = max(C_a, d)$, that is the maximum probability that any given node in community C_a will encounter the destination node d. (6) P_{C_b}.
Output: Message m is delivered to the node b if $b = d$, or $b \in C_a$, or $P_{C_b,d} > P_{C_a,d}$.
Setup: Node a has a message m whose destination is node d.
Events and Associated Actions:

node a encounters a node b

```
1  if $b = d$
2     then node $a$ sends message $m$ to node $b$
3     else
4           if $b \in C_a$
5              then node $a$ sends a copy of
                        the message $m$ to node $b$
6     else
7           if $P_{C_b,d} > P_{C_a,d}$
8              then node $a$ sends a copy of
                        the message $m$ to node $b$
```

Fig. 2: Protocol: MDTN-3PR

B. Security Analysis: Correctness

With each forwarding of the message, the conventional prediction-based routing strategy delivers a copy of the message to a node that has a higher probability of encountering the destination node. We consider our protocol 3PR to be correct if it achieves the same effect as the conventional prediction-based routing strategy.

In 3PR, a node a in community C_a sends message m to node b in a community C_b if a and b encounter and $P_{C_b,d} > P_{C_a,d}$, i.e., if the maximum probability out of all nodes in C_b is higher than the maximum probability out of all nodes in C_a of encountering the destination node (lines 7 and 8). Node b upon receiving the message m floods the message to all nodes in C_b (lines 4 and 5). In Section II, we stated the assumption that a high probability exists of successful message delivery from any source node in a community to any destination node in the community. Given this assumption, the message m reaches all nodes in C_b with high probability, one of which is the node with the maximum probability $P_{C_b,d}$ of encountering the destination node. As $P_{C_b,d} > P_{C_a,d}$, the protocol succeeds (with high probability) in delivering the message m to a node that has a higher probability of encountering the destination node than the node a.

C. Security Analysis: Privacy

A node a only reveals the maximum of the probabilities of the nodes in its community C_a to an outsider node. The maximum is computed within the community in a privacy preserving manner using the MDTN-Private-Max protocol, thus individual probabilities also remain confidential from the nodes inside the community.

One unavoidable side-effect of the protocol is that the adversary learns that node a's probability of encountering the destination node is no higher than the maximum. However, the adversary can learn whether node a is the one who has the maximum, no better than a random guess with probability $1/k$, where k is the number of honest nodes in C_a, and $k \leq n = |C_a|$.

The reader may refer to Sections IV and V for the security analyses of the protocols MDTN-Private-Max and MDTN-Private-Sum respectively.

IV. Privacy Preserving Computation of Maximum

A. Protocol Description

We describe a protocol for computing the maximum of private inputs of nodes in a mobile delay tolerant network. Each node in a set C submits a private number to the protocol and learns the maximum of the private numbers of all nodes in C. The protocol is specified in Figure 3. The protocol uses MDTN-Private-Sum presented in Section V.

Our protocols for private computation of maximum and sum are inspired by the protocols by Kreitz et al. [15], Sheikh and Mishra [16], and Hasan et al. [17], [18]. However, our protocols address specific challenges in MDTNs listed below that the protocols by Kreitz et al. and Sheikh and Mishra do not. Moreover, unlike the protocol by Sheikh and Mishra, our protocols do not require Trusted Third Parties (TTPs).

The mobile delay tolerant network environment presents the following challenges: (1) Mobility implies that the nodes a node will encounter (neighbor nodes) are not known beforehand. (2) Connectivity is intermittent, messages arrive after long and variable delays, and message transmission is asynchronous.

To compute the maximum, each node represents its private number in binary form (Fig. 3: MAXINIT: lines 2 – 5). The protocol proceeds by computing the sum of the most significant bits of the private numbers (MAXROUND: lines 5 and 11). A node knows that it does not have the maximum if its most significant bit was zero but the sum is not zero (MAXROUND: lines 7 and 8). Such a node submits zero to all subsequent sums of bits (MAXROUND: lines 9 and 10). This process continues until the sum of the least significant bit is computed. The nodes compute the sum of the private numbers twice, first with the private numbers of all nodes (MAXROUND: line 3), and second with the private numbers of only the nodes that do not have the maximum (MAXROUND: line 16). The maximum value is obtained as the difference of these two sums (MAXROUND: line 18).

In the given version of the protocol, we assume that only one of the nodes has the maximum value. However, the protocol can be easily extended with an additional round of sum that counts the number of nodes that have the maximum value.

B. Security Analysis: Correctness

The MDTN-Private-Max protocol is composed of $\beta + 2$ rounds of the MDTN-Private-Sum protocol. These $\beta + 2$ rounds are numbered as $j = 0$ to $\beta + 1$ (protocol initiation: lines 2 – 5).

In round $j = 0$, the nodes in C compute $r_0 \leftarrow \sum_{a_u \in C} p_u$ (MAXROUND: lines 2 and 3). Rounds $j = 1$ to β allow each node a_i in C to determine whether it's private value p_i is the maximum of all private values (MAXROUND: lines 4 – 11). In round $j = \beta + 1$, the nodes in C compute $r_{\beta+1} \leftarrow \sum_{a_u \in C} p_u$, where a_u submits $p_u = 0$, if p_u is the maximum (MAXROUND: lines 12 – 17).

Let's assume that the node a_i with the maximum private value p_i correctly determines that it has the maximum. Moreover, let's assume that all other nodes correctly determine that they do not have the maximum. This implies that only the correct a_i submits $p_i = 0$ in round $j = \beta + 1$. Then, all nodes learn the correct maximum when they compute $max \leftarrow r_0 - r_{\beta+1}$ in MAXROUND: line 18, because $r_{\beta+1}$ equals r_0 minus the maximum private value.

We now analyze whether nodes correctly learn if they have the maximum value or not from rounds $j = 1$ to β. For simplicity, let's first assume that in MAXINIT: line 4, $b_{ij} \leftarrow 1$. In round $j = 1$, the nodes in C compute the sum of the most significant bit of binary representations of their private values (MAXROUND: lines 4 and 5). In round $j = 2$, each node a_i reviews the sum of the bits in the previous round (MAXROUND: lines 6 and 7). If the sum is 0 but a_i's input bit was not 0, then it learns that it does not have the maximum value and sets *ismax* ← *false*. This knowledge is correct because even if one other node has 1 as the most significant bit, it implies that the other node's private value is higher than that of a_i. A node who learns that it does not have the maximum, submits 0 as its input bit to all other rounds up to round β (MAXROUND: lines 8 – 10). This process continues until round β. If *ismax* = *true* in round $\beta + 1$ for a node a_i, it implies that in each of the rounds $j = 1$ to β, node a_i submitted 1 as its bit when the sum of the bits was not 0. Thus, assuming that there is only one node with the maximum value, and *ismax* = *true* for node a_i after rounds $j = 1$ to β, then node a_i correctly learns that it has the maximum value.

We assumed that in MAXINIT: line 4, $b_{ij} \leftarrow 1$. This assumption does not change the correctness of the protocol because submitting a random number from a large interval instead of 1 has the same effect (as long as the random number is not 0). Nodes only need to know whether the sum in each round is zero or non-zero.

Protocol: MDTN-Private-Max
Participants: Nodes in a community denoted by the set C. One node in C is the leader node denoted by l.
Input: Each node a_i has a private input p_i.
Output: The nodes in C learn max, which is the maximum of p_i values, where $a_i \in C$.
Setup: (l, g) uniquely identifies an instance of the protocol, where g is an integer. β is the number of bits needed to represent p_i. $msb(p, j)$ is a function that returns the j^{th} most significant bit of a binary representation of p. Nodes are not ordered, that is, a_i denotes any given node in C.
Events and Associated Actions:

leader node l initiates the protocol

```
1   $l$ floods $\langle$MAXINIT$, l, g\rangle$ to all nodes in $C$
2   for $j \leftarrow 0$ to $\beta + 1$
3       do $h \leftarrow g + j$
4           $l$ floods $\langle$MAXROUND$, l, h\rangle$ to all nodes in $C$
5           $l$ waits for completion of
                MDTN-Private-Sum ID:$(l, h)$
```

node a_i receives $\langle$MAXINIT$, l, g\rangle$ from l

```
1   $ismax \leftarrow true$
2   for $j \leftarrow 1$ to $\beta$
3       do if msb$(p_i, j) = 1$
4           then $b_{ij} \leftarrow$ random number
5           else $b_{ij} \leftarrow 0$
```

node a_i receives $\langle$MAXROUND$, l, h\rangle$ from l

```
1   $j \leftarrow h - g$
2   if $j = 0$
3     then $a_i$ participates in MDTN-Private-Sum
              ID:$(l, h)$ to compute $r_0 \leftarrow \sum_{a_u \in C} p_u$
4     else if $j = 1$
5         then $a_i$ participates in MDTN-Private-Sum
              ID:$(l, h)$ to compute $r_1 \leftarrow \sum_{a_u \in C} b_{u1}$
6     else if $2 \leq j \leq \beta$
7         then if $ismax = true$ AND $b_{i(j-1)} = 0$
                  AND $r_{j-1} \neq 0$
8             then $ismax \leftarrow false$
9                 for $y \leftarrow j$ to $\beta$
10                    do $b_{iy} \leftarrow 0$
11            $a_i$ participates in MDTN-Private-Sum
              ID:$(l, h)$ to compute $r_j \leftarrow \sum_{a_u \in C} b_{uj}$
12    else if $j = \beta + 1$
13        then $p_{temp} \leftarrow p_i$
14            if $ismax = true$
15              then $p_i \leftarrow 0$
16            $a_i$ participates in MDTN-Private-Sum
                to compute $r_{\beta+1} \leftarrow \sum_{a_u \in C} p_u$
17            $p_i \leftarrow p_{temp}$
18            $max \leftarrow r_0 - r_{\beta+1}$
```

Fig. 3: Protocol: MDTN-Private-Max

MDTN-Private-Max is composed of multiple rounds of MDTN-Private-Sum. Please refer to Section V-B for a discussion on how MDTN-Private-Sum addresses the specific challenges of Mobile DTN environments.

C. Security Analysis: Privacy

Let's consider a node $a_i \in C$. In an ideal protocol, the node would submit its private value p_i to the TTP. The TTP would compute the maximum of all submitted private values of the nodes in C and disclose only the maximum.

An extra element of information that the MDTN-Private-Max protocol reveals than the ideal protocol is the sum of the private values of the nodes in community C. This is computed at round $j = 0$ in MAXROUND lines 2 and 3. However, the sum is computed using the privacy preserving MDTN-Private-Sum protocol, which does not disclose any individual private values (under the conditions discussed in Section V-C).

In each round $j = 1$ to β (MAXROUND: lines 4 – 11), node a_i submits a random number if the j^{th} bit of its private value is 1 or it submits 0 otherwise. Each round j reveals the sum of the submitted values computed using the MDTN-Private-Sum protocol. Let's assume that the random numbers are non-zero positive real numbers uniformly distributed over a large interval. We discuss three possible cases in terms of the information that can be gleaned in each of these rounds.

Case 1: At least 2 honest nodes have 1 as their j^{th} bit. The adversary (the set of dishonest nodes) can learn whether honest node a_i's j^{th} bit is 1 no better than a random guess with probability $1/k$. The reason is that the only information the sum of random numbers gives to the adversary is that there is at least one node who has 1 as the j^{th} bit.

Case 2: Only one honest node a_x submits a random number and all others submit 0. The node a_x will learn that all other nodes, including a_i, submitted 0. This information tells node a_x that he has the maximum value, however, this is also the output of the protocol. As in the first case, the adversary can learn whether honest node a_i's j^{th} bit is 1 no better than a random guess with probability $1/k$.

Case 3: No node has 1 as the j^{th} bit and thus all nodes submit 0. All nodes will learn that everybody submitted 0. The information that this reveals is that no node has a value higher than $2^{\beta} - 2^{j} - 1$. This information is also revealed from the output of the protocol.

V. Privacy Preserving Computation of Sum

A. Protocol Description

We describe a protocol for computing the sum of private inputs of nodes in a mobile delay tolerant network. Each node in a set C submits a private number to the protocol and learns the sum of the private numbers of all nodes in C. The protocol is specified in Figure 4.

The protocol is initiated by the leader node of a community given as the set of nodes C. The leader node floods an *init* message (Fig. 4: protocol initiation: line 3) to all nodes. After a node receives the *init* message, it sends and receives a random number from each node belonging to C that it encounters (SUMINIT: lines 5 and 6). A node can send the *init* message to an encountered node if it has not received it yet (SUMINIT: lines 3 and 4). After a node has encountered k nodes (SUMINIT: lines 1 and 2), the node sends a partial sum to the leader node (SUMINIT: line 8). A node computes the partial sum as the sum of its private number and all random numbers received minus the sum of all random numbers sent (SUMINIT: line 7). The leader node maintains a running sum of all partial sums received (SUMPARTIAL: line 2). When the partial sums are received from all nodes in C (SUMPARTIAL: line 3), the leader node computes the final sum and floods it to all nodes (SUMPARTIAL: line 4). The final sum is the required sum of the private numbers.

Protocol: MDTN-Private-Sum
Participants: Nodes in a community denoted by the set C. One node in C is the leader node denoted by l.
Input: Each node a_i has a private input p_i.
Output: The nodes in C learn $\sum_{a_i \in C} p_i$.
Setup: (l, g) uniquely identifies an instance of the protocol, where g is an integer. k is a constant such that $2 \leq k < n$, where $n = |C|$. Nodes are not ordered, that is, a_i denotes any given node in C.
Events and Associated Actions:

leader node l initiates the protocol

```
1  $R \leftarrow \phi$
2  $\sigma_C \leftarrow 0$
3  $l$ floods $\langle$SUMINIT$, l, g\rangle$ to all nodes in $C$
```

node $a_i \in C$ receives $\langle$SUMINIT$, l, g\rangle$

```
1  for $j \leftarrow 1$ to $k$
2       do $a_i$ encounters node $a_j \in C$
3          if $a_j$ has not received $\langle$SUMINIT$, l, g\rangle$
4            then $a_i$ sends $\langle$SUMINIT$, l, g\rangle$ to $a_j$
5          $a_i$ sends a random number $r_{ij}$ to $a_j$
6          $a_i$ receives a random number $r_{ji}$ from $a_j$
7  $\sigma_i \leftarrow p_i - \sum_{j=1}^{k} r_{ij} + \sum_{j=1}^{k} r_{ji}$
8  $a_i$ sends $\langle$SUMPARTIAL$, l, g, \sigma_i\rangle$ to $l$
```

leader node l receives $\langle$SUMPARTIAL$, l, g, \sigma_i\rangle$ from a_i

```
1  $R \leftarrow R \cup \{a_i\}$
2  $\sigma_C \leftarrow \sigma_C + \sigma_i$
3  if $R = C$
4     then $l$ floods $\langle$SUMFINAL$, l, g, \sigma_C\rangle$ to all nodes in $C$
```

Fig. 4: Protocol: MDTN-Private-Sum

B. Security Analysis: Correctness

The first challenge for the protocol due to the mobile delay tolerant network environment is that the nodes a node will encounter (neighbor nodes) are not known beforehand. To address this challenge, the protocol allows a node $a_i \in C$ to encounter any other k nodes in C (SUMINIT: lines 1 and 2). The encountered nodes, given as a_j, where $j \in \{1, 2, \ldots, k\}$, are considered as the neighbors of node a_i.

Each node $a_i \in C$ sends a random number r_{ij} to each encountered node a_j (SUMINIT: lines 5 and 6). Node a_i adds $-r_{ij}$ to its sum σ_i, whereas node a_j adds r_{ij} to its sum σ_j (SUMINIT: line 7). Each node a_i also adds its private value p_i to its sum σ_i (SUMINIT: line 7). When the leader node computes $\sigma_C = \sum_{a_i \in C} \sigma_i$, the sum σ_C is the required sum $\sum_{a_i \in C} p_i$ because $-r_{ij}$ and r_{ij} added to σ_i and σ_j

respectively add up to the identity 0 (SUMPARTIAL: lines 1 – 4).

The second set of related challenges of mobile delay tolerant network environments are as follows: connectivity is intermittent, messages arrive after long and variable delays, and message transmission is asynchronous. The following two elements of the protocol address this set of challenges: (1) The *init* message (SUMINIT) reaches all nodes in C with high probability and thus they all participate in the protocol. This is due to the assumption that a high probability exists of successful message delivery from any source node to any destination node in a community. (2) If a node $a_i \in C$ that has received the *init* message encounters a node $a_j \in C$ that has not yet received the *init* message then a_i sends a copy of the message to a_j to initiate it to the protocol (SUMINIT: lines 3 and 4). Nodes consider an encounter successful only if they exchange all messages according to the specification during their period of contact. Otherwise, they ignore any partial messages sent and received.

C. *Security Analysis: Privacy*

Let's consider a node $a_i \in C$. In an ideal protocol, the node would submit its private value p_i to a TTP. The TTP is considered trustworthy therefore it would not disclose the private value p_i of node a_i to any other party. It would only reveal the output of the protocol, which is the sum of the private values received from all nodes in C.

In the MDTN-Private-Sum protocol, node a_i discloses the following information: (1) One random number to each of the k nodes that it encounters after receiving the SUMINIT message. (2) The value σ_i to the leader node l as part of the SUMPARTIAL message. The value σ_i is also revealed to the intermediate nodes that participate in the delivery of the message to the leader node.

The random numbers r_{ij}, where $j \in \{1, 2, \ldots, k\}$, are independent of p_i therefore they reveal no info about p_i.

$\sigma_i = p_i + \gamma_i$, where $\gamma_i = -\sum_{j=1}^{k} r_{ij} + \sum_{j=1}^{k} r_{ji}$. Let's assume that the interval of the random numbers is large compared to the interval of p_i and that the random numbers are distributed uniformly. This implies that the interval of γ_i is also large and that it is distributed uniformly. Thus there is high probability that the adversary can learn no information about p_i from σ_i.

The adversary can learn p_i if it learns γ_i in addition to σ_i. To learn γ_i, the adversary must learn all values r_{ij} and r_{ji}. This is possible only if all k nodes a_j that encountered node a_i are dishonest and collude to reveal all of their individual r_{ij} and r_{ji} values and consequently the value of γ_i.

As in the ideal protocol, the output of the protocol is the sum of the private values of all nodes in C. The MDTN-Private-Sum protocol thus does not reveal any more information about the private value p_i of node a_i than the ideal protocol if the following assumptions hold true: (1) the interval of the random numbers r_{ij} and r_{ji} is large compared to the interval of p_i and the random numbers are distributed uniformly, and (2) at least one of the k nodes that encountered node a_i is honest.

VI. Performance Evaluation

In this section, we present the performance evaluation of 3PR. We start by presenting our simulation settings and the mobility model in sections VI-A and VI-B respectively. We then introduce the routing protocols against which we compare the performance of 3PR in section VI-C. Finally, we present the results of our experiment in Section VI-D. As none of the non-naive algorithms against which we compare 3PR are privacy preserving, the objective of this performance evaluation is to assess the cost of introducing privacy preservation mechanisms in the routing process.

A. *Simulation Settings*

We have implemented 3PR as a module of the The Opportunistic Network Environment simulator, i.e., The ONE [19].

We have used a simulation area of 3000×1500 m^2. This area is equally divided into 12 regions. In each region, we initially deploy a varying number of nodes (from ten to fifty). Each node considers the region in which it has been deployed as its *local region*. According to the mobility model we used, further described below, a node is more likely to visit its local region than other places. Nodes associated to a region constitute a community. This simulation scenario is very similar to the one used in PRoPHET [6].

The transmission range and rate are set as 10m and 2Mb/s respectively. This is consistent with contemporary protocols such as Bluetooth [20]. The speed of nodes is set to 1.34m/s, which is an average human walking speed [21]. Each experiment simulates thirteen hours. The first hour is a warm up period during which no message is generated. After this period, every thirty seconds, a random source node sends a message to a random destination node. We have considered only messages for which the source and the destination belong to different communities.

B. *Mobility Model*

In our evaluation, we adopt the community-based mobility model proposed in [11], which has been widely utilized for the evaluation of community-based routing protocols [22][23][12]. In this mobility model, each community is associated with a geographical area. The movement of node i, which belongs to the community C_i consists of a sequence of *local* and *roaming* epochs. A local epoch is a random direction movement restricted inside the area associated with the community C_i. A roaming epoch is a random direction movement inside the entire network. If the previous epoch of node i was a local one, the next epoch is a local one with probability p_l, or a roaming epoch with probability $1 - p_l$. Similarly, if the previous epoch of node i was a roaming one, the next epoch is a roaming one with probability p_r, or a local one with probability $1 - p_r$. The state transition between local and roaming epochs is shown in Figure 5. In our simulations, we adopt the same values for p_l and p_r as in [6], i.e., p_l=0.8 and p_r=0.2.

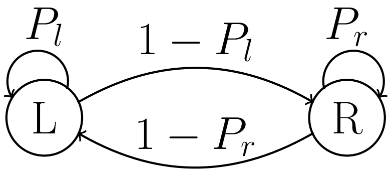


Fig. 5: Community-based Mobility Model

C. Benchmark

We have compared the performance of 3PR against the following protocols:

Epidemic: In this protocol, a node forwards a copy of each unexpired message it holds to every node it encounters, which does not already have a copy of the message. The protocol delivers the upper bound on delivery ratio.

Direct: In this protocol, the source node only forwards the message to the destination node. The protocol delivers the lower bound on delivery ratio.

PRoPHET: In this prediction-based protocol, a node forwards a copy of a message it holds to a node it encounters, only if the latter has a higher probability of encountering the destination node of the message. The parameters of the protocol are set as described in [6].

Bubble: This community-based protocol utilizes social information about nodes, such as their centrality and the community to which they belong. In this protocol, a message is forwarded based on the global rankings of two encountering nodes, until it reaches a node in the same community as the destination node. After that, the message is forwarded based on the local rankings of two encountering nodes, until it either reaches the destination node or expires.

D. Performance Results

We use the following three well know metrics: (1) Delivery ratio – the ratio of messages that have been delivered to the total unique messages created; (2) Delivery cost – the total number of messages transmitted in the simulation, normalized by dividing by the total number of unique messages created; and (3) Average Delivery latency – the average time needed to finish transmitting messages to their destinations.

Figure 6a shows the delivery ratio of the compared protocols as a function of the time to live (TTL) of the generated messages. We observe that Epidemic and Direct achieve the best and the worst delivery ratio, respectively, for all values of TTL. We also observe that PRoPHET achieves a better delivery ratio than 3PR overall. Nevertheless, the delivery ratio of 3PR approaches the one of PRoPHET when the TTL is greater or equal to 4 hours. Indeed, the difference between the performance of the two protocols becomes lower than 10%. Finally, 3PR has a higher delivery ratio than Bubble. The difference between the performance of the two protocols increases up to 40% for a TTL of 3 hours.

Figure 6b, shows the delivery cost of the compared routing protocols. We observe that Epidemic and Direct have the highest and lowest delivery cost, respectively, whatever the value of TTL. Compared to the others, Bubble has a low delivery cost, which remains stable as the TTL increases. 3PR has a delivery cost that is slightly higher than the one of Bubble (around 30% higher on average), but that is much lower than the one of PRoPHET (an order of magnitude lower on average).

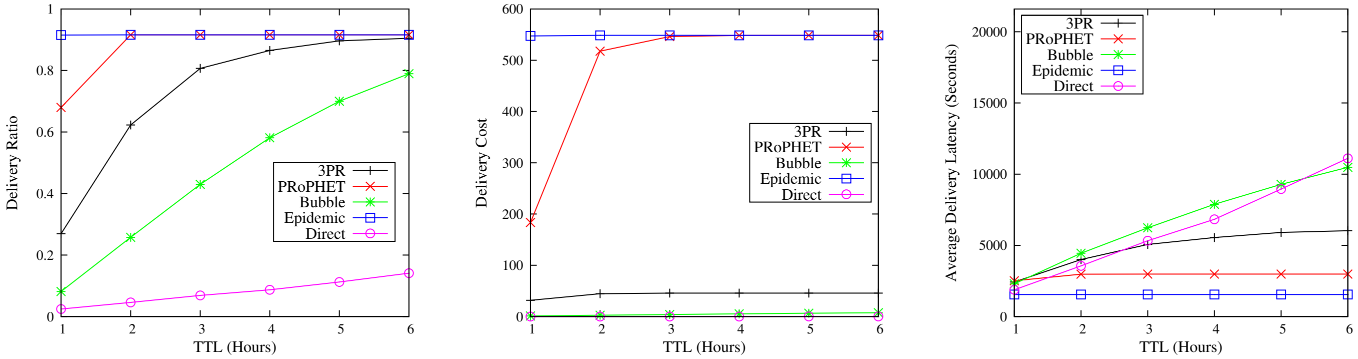
Figure 6c shows the delivery latency of the compared routing protocols. We observe that Epidemic has the lowest delivery latency, whatever the TTL. Further, PRoPHET follows the same trend as Epidemic with higher latencies (around 1500 seconds higher). The delivery latency of 3PR gradually increases as the TTL increases. The performance of Bubble and Direct increases linearly with the increase of the TTL. We observe that 3PR delivers messages with a lower latency on average compared to Bubble and Direct.

VII. Related Work

The existing body of work on privacy preserving routing in DTNs can be broadly divided into two categories: (1) protocols that preserve the privacy of nodes in the context of their identity and location, and (2) protocols that preserve the confidentiality of the content of routed messages. In contrast, our protocol 3PR is a novel type of protocol which has the specific goal of hiding the encounter probabilities of nodes. Therefore, 3PR differs fundamentally from other existing privacy preserving routing protocols for DTNs due to the difference in objectives. We refer the reader to [8] for a detailed classification of privacy preserving routing protocols.

We note some recent protocols that attempt to hide the identity and location of nodes. Lu et al. [24] propose the SPRING protocol for VANETs which prevents the adversary from analyzing packets to find out the identity of the source and destination nodes. Lu et al. [24] present the ALAR routing protocol for MANETs which hides a sender's location by fragmenting a message and forwarding each segment to different receivers. Defrawy and Tsudik [25] present the PRISM routing protocol for MANETs which is resistant to attacks that aim to track the location of nodes. Kate et al. [26] present an anonymous communication architecture for DTNs using Identity-Based Cryptography (IBC) in which the identity of the source node is removed by trusted gateways.

The following works protect the confidentiality of messages in DTNs. Jansen et al. [27] propose a protocol in which a message is divided into multiple shares using secret sharing. The shares are delivered to the destination via multiple independent paths thus protecting the content of the message. Shi et al. [28] propose ARDEN, a privacy-preserving scheme based on onion routing. Instead of the keys of individual intermediate nodes, ARDEN encrypts messages with the keys of social groups. Nodes in the same group share the same key, therefore they can all participate in message forwarding thus improving the probability of delivery.



(a) Delivery ratio wrt the increasing TTL of messages.

(b) Delivery cost wrt the increasing TTL of messages.

(c) Delivery latency wrt the increasing TTL of messages.

Fig. 6: Performance comparison of the routing protocols.

VIII. Conclusion

In this paper, we presented 3PR, the first privacy-preserving prediction-based routing protocol for Mobile DTNs. 3PR takes advantage of the mobility patterns of nodes to route messages, yet preserves the privacy of nodes by hiding their individual mobility patterns. The protocol requires that nodes in a community compute the maximum probability that a node in the community will encounter a destination node. We presented a protocol that computes this maximum in mobile delay tolerant networks in such a manner that the individual private values are not revealed even to the nodes inside the community. We evaluated 3PR both theoretically, with correctness and privacy analyses, and practically, through extensive simulations. Our simulations on a well established community-based mobility model, demonstrate that 3PR has comparable performance to existing prediction-based protocols, while being the only one that preserves the privacy of nodes.

References

[1] K. Fall, "A delay-tolerant network architecture for challenged internets," in *Proc. of the conf. on applications, technologies, architectures, and protocols for computer communications*, 2003, pp. 27–34.

[2] M. Liu, Y. Yang, and Z. Qin, "A survey of routing protocols and simulations in dtns," in *Proc. of WASA*, 2011, pp. 243–253.

[3] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of human mobility on opportunistic forwarding algorithms," *IEEE Transactions on Mobile Computing*, pp. 606–620, 2007.

[4] P. Hui, J. Crowcroft, and E. Yoneki, "Bubble rap: Social-based forwarding in delay-tolerant networks," *IEEE Trans. Mobile Comput.*, vol. 10, no. 11, pp. 1576–1589, 2011.

[5] A. Mashhadi, S. Ben Mokhtar, and L. Capra, "Habit: Leveraging human mobility and social network for efficient content dissemination in delay tolerant networks," in *Proc. of WoWMoM 2009*, 2009, pp. 1–6.

[6] A. Lindgren, A. Doria, and O. Schelén, "Probabilistic routing in intermittently connected networks," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 7, no. 3, pp. 19–20, 2003.

[7] S. Gambs, M.-O. Killijian, and M. N. n. del Prado Cortez, "Show me how you move and i will tell you who you are," in *Proc. of the ACM SPRINGL 2010*, 2010, pp. 34–41.

[8] J. Miao, O. Hasan, S. B. Mokhtar, L. Brunie, and K. Yim, "An investigation on the unwillingness of nodes to participate in mobile delay tolerant network routing," *International Journal of Information Management (Elsevier)*, http://dx.doi.org/10.1016/j.ijinfomgt.2012.11.001, 2013.

[9] C. Boldrini and A. Passarella, "Hcmm: Modelling spatial and temporal properties of human mobility driven by users' social relationships," *Computer Communications*, vol. 33, no. 9, pp. 1056 – 1074, 2010.

[10] A. Vahdat, D. Becker *et al.*, "Epidemic routing for partially connected ad hoc networks," CS-200006, Duke University, Tech. Rep., 2000.

[11] T. Spyropoulos, K. Psounis, and C. Raghavendra, "Performance analysis of mobility-assisted routing," in *Proc. of the 7th ACM Intl. Symp. on Mobile ad hoc networking and computing*, 2006, pp. 49–60.

[12] H. Dang and H. Wu, "Clustering and cluster-based routing protocol for delay-tolerant mobile networks," *Wireless Communications, IEEE Transactions on*, vol. 9, no. 6, pp. 1874–1881, 2010.

[13] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Efficient routing in intermittently connected mobile networks: The single-copy case," *IEEE/ACM Trans. Netw.*, vol. 16, no. 1, pp. 63–76, 2008.

[14] O. Goldreich, *The Foundations of Cryptography - Volume 2*. Cambridge University Press, 2004.

[15] G. Kreitz, M. Dam, and D. Wikstrom, "Practical private information aggregation in large networks," in *Proceedings of Nordsec*, 2010.

[16] R. Sheikh and D. K. Mishra, "Protocols for getting maximum value for multi-party computations," in *Proc. of the 4th Asia Intl. Conf. on Mathematical/Analytical Modelling and Computer Simulation*, 2010.

[17] O. Hasan, L. Brunie, and E. Bertino, "Preserving privacy of feedback providers in decentralized reputation systems," *Computers & Security*, vol. 31, no. 7, pp. 816 – 826, October 2012.

[18] O. Hasan, L. Brunie, E. Bertino, and N. Shang, "A decentralized privacy preserving reputation protocol for the malicious adversarial model," University of Lyon, INSA-Lyon, LIRIS, France, Tech. Rep. RR-LIRIS-2012-008, June 2012.

[19] A. Keränen, J. Ott, and T. Kärkkäinen, "The one simulator for dtn protocol evaluation," in *Proc. of the 2nd Intl. Conf. on Simulation Tools and Techniques*, 2009.

[20] A. Keränen and J. Ott, "Increasing reality for dtn protocol simulations," *Helsinki University of Technology, Tech. Rep., July*, 2007.

[21] M. Kim, D. Kotz, and S. Kim, "Extracting a mobility model from real user traces," in *Proc. of INFOCOM 2006*, 2006, pp. 1–13.

[22] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Efficient routing in intermittently connected mobile networks: The multiple-copy case," *IEEE/ACM Trans. Netw.*, vol. 16, no. 1, pp. 77–90, 2008.

[23] ——, "Spray and focus: Efficient mobility-assisted routing for heterogeneous and correlated mobility," in *Proc. of PerCom Workshops*, 2007, pp. 79–85.

[24] X. Lu, P. Hui, D. Towsley, J. Pu, and Z. Xiong, "Anti-localization anonymous routing for delay tolerant network," *Computer Networks*, vol. 54, no. 11, pp. 1899 – 1910, 2010.

[25] K. Defrawy and G. Tsudik, "Privacy-preserving location-based on-demand routing in manets," *Selected Areas in Communications, IEEE Journal on*, vol. 29, no. 10, pp. 1926–1934, 2011.

[26] A. Kate, G. Zaverucha, and U. Hengartner, "Anonymity and security in dtns," in *Proc. of SecureComm*, 2007, pp. 504–513.

[27] R. Jansen and R. Beverly, "Toward anonymity in dtns: Threshold pivot scheme," in *Proc. of MILCOM*, 2010, pp. 587–592.

[28] C. Shi, X. Luo, P. Traynor, M. H. Ammar, and E. W. Zegura, "Arden: Anonymous networking in delay tolerant networks," *Ad Hoc Networks*, vol. 10, no. 6, pp. 918–930, 2012.