PRIVO: A PRIvacy-preserVing Opportunistic routing protocol for Delay Tolerant Networks

Naercio Magaia, Carlos Borrego, Paulo Pereira and Miguel Correia

***Abstract*—This paper introduces PRIVO, a PRIvacy-preserVing Opportunistic routing protocol for Delay Tolerant Networks. PRIVO models a DTN as a time-varying neighboring graph where edges correspond to the neighboring relationship among pairs of nodes. PRIVO ensures privacy by protecting each node's sensitive information even if it has to be processed elsewhere. In addition, nodes also compare their routing metrics in a private manner using the Paillier homomorphic encryption scheme.**

The effectiveness of PRIVO is supported through extensive simulations with synthetic mobility models and real mobility traces. Simulations results show that PRIVO presents on average cryptographic costs below 1% in most scenarios. Additionally, PRIVO presents on average gains of 22.2% and 39.7% in terms of delivery ratio for the synthetic and real scenarios considered, respectively.

***Index Terms*—Privacy, Routing, Delay Tolerant Networks, Betweenness centrality, Similarity**

I. Introduction

DELAY Tolerant Networks (DTNs) [1] are networks in which end-to-end connectivity between a source and target node may never exist. DTN nodes rely on opportunistic routing where a store-carry-and-forward approach is used, that is, DTN nodes store (or buffer) messages and forward them to others until they reach their target.

DTN routing involves the challenging task of finding suitable nodes to forward messages to. To address this problem, static and dynamic network information has been used [2]. Through social network analysis, static network information, which is more stable over time, can be leveraged and used by DTN routing protocols to facilitate the forwarding of messages. Centrality [3], which is widely used in graph theory and network analysis, is a quantitative measure of the structural importance of a certain node in relation to others within the network. In DTNs, central nodes may be considered good candidates to be relay nodes. Among the centrality metrics, betweenness centrality [3] can be considered the most prominent, as it measures how well a node can facilitate communication among others by summing up the fraction of shortest paths between other pairs of nodes passing through it. Similarity [3], which is a measure of common features of a group of nodes, can be computed, for example, by finding common neighbor nodes they might have. Computing routing metrics such as betweenness centrality or similarity requires the exchange of information between nodes.

In a DTN, nodes represent individuals, vehicles, or other entities, and edges the relationship between two entities. In DTNs there is information that may be private, such as the entities owning and managing DTN nodes and their relationships. DTN applications would benefit from mechanisms that enforce the entities' identities and/or relationship anonymity due to the sensitive or confidential nature of the entities' identities and their behaviors. A DTN node may disclose private information by sending private data to other nodes. Privacy-preservation techniques allow protecting privacy through masking, modification and/or generalization of the original data without sacrificing the data utility.

If it is considered that some nodes might misbehave [4], private information such as contacts' history, list of neighbors, etc., which is required for computing some routing metrics should not be disclosed to misbehaving nodes. However, nodes should be able to use part of this information, if necessary. Furthermore, despite the good routing performance of some of the proposed routing protocols [2], [3], most of the security issues presented in [4] (such as confidentiality, integrity, privacy, etc.) were not considered. For instance, to deal with confidentiality and privacy, nodes should implement cryptographic protocols.

In this article, a PRIvacy-preserVing Opportunistic routing protocol for Delay Tolerant Networks (PRIVO) is presented. PRIVO models a DTN as a time-varying neighboring graph where the edges correspond to the neighboring relationship among pairs of nodes. It ensures privacy by protecting each node's sensitive information even if it has to be processed elsewhere.

The contributions of this paper are summarized as follows:

- PRIVO, an efficient privacy-preserving routing protocol for DTNs;
- PRIVO weight (*pweight*), which is a time-varying metric based on nodes' encounter history, is defined in order to assess the neighboring relationship among pairs of

This work was partially supported by Fundação Calouste Gulbenkian and by national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UID/CEC/50021/2013.

Naercio Magaia, Paulo Pereira and Miguel Correia are with INESC-ID, Instituto Superior Técnico, Universidade de Lisboa, Rua Alves Redol, nº 9, 1000-029 Lisboa, Portugal. (phone: +351-213100286; fax: +351-213145843; e-mails: naercio.magaia@tecnico.ulisboa.pt, prbp@inesc.pt, miguel.p.correia@tecnico.ulisboa.pt).

Carlos Borrego is with the Department of Information and Communications Engineering, Autonomous University of Barcelona, Barcelona, Spain (email: cborrego@deic.uab.cat)

ISBN 978-3-901882-94-4

nodes;
- Two anonymization methods (i.e., binary anonymization and neighborhood randomization) to ensure privacy are defined. They are used by DTN nodes to exchange neighborhood information;
- A privacy mechanism that uses the Paillier homomorphic encryption scheme, to allow nodes to compare their routing metrics without disclosing them, is proposed.

The remainder of this paper is organized as follows. Section II presents related work. Section III introduces relevant social metrics and cryptographic mechanisms for this work. Section IV presents the PRIVO protocol. In Sections V and VI, simulation models and results are presented. Finally, Section VII presents concluding remarks.

II. Related work

According to the literature [5], privacy breaches can be classified as identity disclosure, link disclosure and attribute disclosure. Identity disclosure is the case when the identity of the individual associated to the node is revealed. Link disclosure happens when the sensitive relationship between the individuals is disclosed. Attribute disclosure is the case when the sensitive data associated with the node is compromised. Moreover, there are several types of sensitive information such as node attributes, specific link relationships between nodes, nodes degrees, neighborhoods of some target nodes, etc.

Anonymization methods [5] can be used to protect the privacy of information if sensitive information needs to be processed elsewhere. There are three main anonymization methods, namely: (*i*) *k*-anonymity privacy preservation via edge modification, that modifies graph structure by successive deletions and additions of edges so that each node in the modified graph is indistinguishable with at least $k-1$ other nodes in terms of a given network property; (*ii*) edge randomization, that modifies the graph structure by randomly adding/deleting edges or by switching edges; and (*iii*) cluster-based generalization, where nodes and edges are clustered into groups and anonymized into a super-node.

It is commonly assumed by DTN routing protocols [2], [3] that nodes are willing to share their private information for the sake of the network's performance. Some routing protocols that address privacy issues in DTNs have been proposed [6]–[11]. Routing approaches such as [6]–[8] ensure attribute privacy. The location used by the source node to send messages is protected in [6]. The context, e.g. personal information, residence, work, hobbies, interest profiles, etc., which is used for forwarding is protected in [7], [8]. In [9], an adaptive mechanism for achieving user anonymity that ensures identity privacy is proposed. Identity privacy can be compromised if an attacker combines external knowledge with observed network structure [5]. In [10], an approach that ensures link privacy has been proposed where instead of transmitting the list of friends of the sender as a list of nodes, a modified and obfuscated one is transmitted.

Other privacy techniques have been proposed in the literature, but may not be adequate for DTNs. For instance, with homomorphic encryption [12] – proposed by Rivest et al. in 1978 – a node can carry out computations on encrypted values, without needing to decrypt them first. In [13] and [11], privacy-preserving routing protocols based on additive homomorphic encryption (Paillier cryptosystem [14]) were proposed. The former, which was proposed for peer-to-peer networks, allowed a node to calculate its similarity to other nodes using multivariate polynomial evaluation, meanwhile the latter, which was proposed as a secure geographical routing protocol for DTNs, allowed nodes to compare their habitats in order to choose the best forwarder for every message, respectively. Besides [11], which is not suitable for social DTNs and only ensures attribute privacy, none of the above protects the nodes' private information if it has to be shared and processed elsewhere (link privacy), or used during routing decisions (attribute privacy). PRIVO ensures both link and attribute privacy.

III. Background

A. Assumptions and notations

A notation similar to [3] is used. A DTN neighboring graph is modeled as a time-varying graph $\mathcal{G} = (V, E, \mathcal{T}, w)$ where each vertex $v \in V$ corresponds to a node in the network and each edge $e = (i, j) \in E$ represents the relationship between these nodes (i.e., that these nodes have encountered before). The relations among nodes are assumed to take place over a time span $\mathcal{T} \in \mathbb{T}$ known as the lifetime of the network; $w: E \times \mathcal{T} \rightarrow [0,1]$ is called *weight* function and indicates the strength of an edge at a given time.

Let a footprint of $\mathcal{G}$ from t_1 to t_2 be defined as a static graph $G^{[t_1,t_2)} = (V, E^{[t_1,t_2)})$ such that $\forall e \in E, e \in E^{[t_1,t_2)} \iff \exists\, t \in [t_1, t_2),\ w(e,t) \in [0,1]$, i.e., the footprint aggregates all interactions of a given time window (or timeslot) into static graphs. Let $\tau = [t_0, t_1), [t_1, t_2), \ldots, [t_i, t_{i+1}), \ldots$ (where $[T_k, T_{k+1})$ can be noted τ_k) be the lifetime $\mathcal{T}$ of the time-varying graph partitioned in sub-intervals. The sequence $SF(\tau) = G^{\tau_0}, G^{\tau_1}, \ldots$ is called sequence of footprints of $\mathcal{G}$ according to τ.

B. Social metrics

A variety of network information has been used to address the challenging task of finding the most suitable node to forward messages in a DTN, namely dynamic network information (e.g., location, traffic, encounter information, etc.) and social network information (e.g., social relations among nodes). However, social network information is more stable over time than dynamic network information and can be leveraged by DTN routing protocols to facilitate the forwarding of messages [2].

1) Ego betweenness centrality

Centrality of a node in a network is a quantitative measure of the structural importance of this node in relation to others within the network. Typically, a node can be considered as central if it plays an important role in the network's connectivity, for example, if it is more apt to connect to others in the network. The three most common centrality metrics are degree, closeness and betweenness centrality [3]. Degree centrality is defined as the number of links (that is, direct neighbors) incident upon a given node. Closeness centrality is defined as the total shortest path distance from a given node to all other nodes. Betweenness

centrality is defined as the number of geodesics (shortest paths) passing through a given node. Betweenness centrality can be perceived as a measure of the load placed on a given node since it measures how well a node can facilitate communication among others. PRIVO uses a betweenness centrality metric that does not require global knowledge, hence being more suitable for DTNs.

An ego network [15] (also known as the neighborhood network of the ego) is defined as a network that consists of a central node (ego) along with its direct neighbors (the other nodes the ego is directly connected to) and all links among these neighbors. The shortest paths, due to the structure of the ego network, are either of length 1 or 2. Every single pair of non-adjacent direct neighbors must have a shortest path of length 2 which passes through the ego. Shortest paths of length 1 do not contribute to the betweenness centrality computation. If $\boldsymbol{\mathcal{A}}$ is an adjacency matrix of graph $\mathcal{G}$, then $\boldsymbol{\mathcal{A}}^2_{i,j}$ contains the number of geodesics of length 2 connecting vertices i and j. The number of shortest paths between i and j is given by $\boldsymbol{\mathcal{A}}^2[\mathbf{1}-\boldsymbol{\mathcal{A}}]_{i,j}$ (where $\mathbf{1}$ is a matrix of all 1's).

The ego betweenness centrality (c_{EBC}) is the sum of the halved reciprocal entries $\boldsymbol{\mathcal{A}}^2[\mathbf{1}-\boldsymbol{\mathcal{A}}]_{i,j}$ such that $\boldsymbol{\mathcal{A}}_{\boldsymbol{i,j}}=\mathbf{0}$.

2) Similarity

Similarity [16] expresses the amount of common features of a group in social networks. In sociology, the probability of two individuals being acquainted increases with the number of common acquaintances between them [17]. In computer networks, similarity between nodes i and j can be defined as the number of common neighbors among them. Therefore, the more common neighbors they have, the more similar they are.

C. Homomorphic encryption

In cryptography, finding common elements in two private sets without exposing the sets themselves is known as the Private Set Intersection (PSI) problem [18]. For instance, an algorithm that solves the PSI problem would allow a trusted node to send an encrypted version of some data to be processed by an untrusted node and the latter would perform computations on this encrypted data without knowing anything of the data's real value, and send back the result. The trusted node would expect the decrypted result to be equal to the intended computed value, as if it was performed on the original data. For example, with homomorphic encryption a node can carry out computations on encrypted values, without decrypting them first.

An additive homomorphic encryption scheme is the one in which two numbers encrypted with the same key $\mathcal{E}(a)$ and $\mathcal{E}(b)$ can be added without being first decrypted, i.e., one can efficiently compute $\mathcal{E}(a+b)$ without decrypting them.

In the Paillier cryptosystem [19], which is an additive homomorphic encryption scheme, when entity i wants to send message m to entity j, entity i selects random primes p and q and constructs $n=pq$; plaintext messages are elements of $\mathbb{Z}_n$ and cyphertext are elements of $\mathbb{Z}_{n^2}$. Entity i picks a random $g\in\mathbb{Z}^*_{n^2}$ and verifies that $\exists\mu$ where $\mu=(L(g^\lambda \bmod n^2))^{-1} \bmod n$, $L(x)=(x-1)/n$ and $\lambda=\mathrm{lcm}(p-1,q-1)$. If $\nexists\mu$ then a new random $g\in\mathbb{Z}^*_{n^2}$ must be picked. Entity i's public key (p_k) is (n,g) and private key (s_k) is (λ,μ).

To encrypt a message m, entity j picks a random $r\in\mathbb{Z}^*_n$ and computes the cypher text $c=\mathcal{E}(m)=g^m\cdot r^n \bmod n^2$, therefore cyphering with p_k. To decrypt c, entity i computes $\mathcal{D}(c)=\left(L(c^\lambda \bmod n^2)\right)^{-1}\cdot\mu \bmod n=m$, therefore deciphering with s_k.

Let $\mathcal{E}(a)=g^a\cdot r_1^n \bmod n^2$ and $\mathcal{E}(b)=g^b\cdot r_2^n \bmod n^2$. Entity j can compute the sum this way: $\mathcal{E}(a+b)=\mathcal{E}(a)\cdot\mathcal{E}(b) \bmod n^2=g^{a+b}\cdot(r_1r_2)^n \bmod n^2$.

Let $\mathcal{E}(a)=g^a\cdot r_1^n \bmod n^2$ and k be a non-encrypted constant. Entity j can compute the multiplication by a non-encrypted constant $\mathcal{E}(k\cdot a)=\mathcal{E}(a)^k \bmod n^2=g^{k\cdot a}\cdot(r_1)^n \bmod n^2$.

IV. The PRIVO protocol

The PRIvacy-preserVing Opportunistic routing protocol for Delay Tolerant Networks (PRIVO) detects and utilizes the inherent social network structure to facilitate packet forwarding in DTNs. It models a DTN as a time-varying neighboring graph where vertices correspond to nodes and edges correspond to the neighboring relationship among pairs of nodes.

PRIVO ensures privacy by means of anonymization and homomorphic encryption. It uses anonymization to avoid disclosing historical information associated to each node's neighboring graph. Moreover, when two nodes meet, they do not share private information associated with their routing metrics, which is necessary to identify the best message forwarder (i.e., the most suitable node to forward a given message). Nodes compare these metrics in a private manner using homomorphic encryption.

The PRIVO protocol is composed of the following steps: construction and anonymization of the neighboring graph, determination of routing metrics and the routing algorithm.

A. Construction of the neighboring graph

Let $x_{i,j}(t)$ denote the separation period between nodes i and j, τ denote the elapsed time and $n_{i,j}$ be the number of times that nodes i and j were away from each other. So, $x_{i,j}(t)=0$ means that nodes i and j are within communication range at time t, otherwise $x_{i,j}(t)=1$. The time-varying average separation period (hereafter average separation period) is given by

$$\delta_{i,j}(x)=\frac{\int_\tau x_{i,j}(t)dt}{n_{i,j}}$$

The normalized average separation period $\hat{\delta}_{i,j}$ is given by

$$\hat{\delta}_{i,j}=1-\frac{\delta_{i,j}}{\tau}$$

and the unbiased variance estimator is given by

$$\hat{\sigma}(x)=\frac{1}{l}\sum_{k=1}^{l}(x_k-\delta(x))^2$$

The average separation period aims at capturing the evolution of social interactions in similar time-periods (or timeslots).

In here, daily timeslots were considered. The average separation period in the same timeslot over consecutive days is updated using an exponential weighted moving average as follows

$$\hat{\delta}^t = (1-\alpha)\cdot\hat{\delta}^{t-1} + \alpha\cdot\hat{\delta}^t$$

where α is the smoothing factor, and $0 < \alpha < 1$, and it is depreciated over consecutive timeslots as follows

$$\hat{\delta}^\tau = (1-\alpha)\cdot\hat{\delta}^{\tau-1}$$

The unbiased variance estimator is updated as follows

$$\hat{\sigma}^t = (1-\beta)\cdot\hat{\sigma}^{t-1} + \beta\cdot\left(x-\hat{\delta}\right)^t$$

The social strength among nodes in a specific daily timeslot may provide insights on their social strength in consecutive timeslots on the same day, therefore increasing the probability of nodes being capable of transmitting data as transmissions could be resumed, with high probability, on the same timeslot on the next day [20]. The time-varying PRIVO weight (hereafter *pweight*), $w_{i,j}$, over a daily timeslot is given by

$$w_{i,j} = \frac{1}{\eta}\sum_{k=1}^{\eta}\hat{\delta}_k$$

where $\eta = |\tau|$ is the number of timeslots (or sub-intervals); *pweight* shows the neighboring relationship among nodes and gives hints about the forwarding opportunities between them, i.e., larger $w_{i,j}$ indicates a better future contact probability between nodes i and j.

In PRIVO, nodes' routines are used to quantify the time-varying strength of social ties between nodes. For instance, if daily routines are considered, each node computes the average separation periods to other nodes during the same set of daily timeslots over consecutive days.

B. Anonymization of the neighboring graph

A DTN node may disclose private information by sending private data to other nodes. Privacy-preservation techniques allow protecting privacy through masking, modification and/or generalization of the original data without sacrificing data utility.

PRIVO deals with link disclosure since each node's ego network contains the list of neighbors and their social strengths. PRIVO proposes two anonymization techniques that are suitable for DTNs as they ensure data utility: neighborhood randomization and binary anonymization.

Neighborhood randomization consists in partially hiding each node's neighboring graph containing its historical encounter information. When two nodes are in communication range, they only exchange the least possible number of nodes in their neighboring graphs. If w_{ij} is high, it might mean that nodes i and j have a strong tie (i.e., that they meet often), or even that they have met recently. The latter may be a random link, i.e., a recent occasional connection that looks like a strong tie.

Neighborhood randomization works as follows: upon an encounter between nodes i and j, each node selects a random number between $\chi \in \mathbb{N}$ and the total number of nodes in its neighboring graph. χ should be selected taking into account the amount of information that each node is willing to share. Note that there is a tradeoff between the amount of information to share and the performance of the routing protocol. Sharing less information might compromise the utility of the randomized neighboring graph. If too much information is shared, the node might be disclosing too much private information. The set of nodes to share between i and j is then randomly selected among all possible ones and it is limited by the smallest previous randomly selected number. This allows hiding each node's degree. Randomly selecting nodes to add to the anonymized neighboring graph that will be shared allows mixing random contacts with strong contacts, therefore hiding the contact patterns among neighbors since *pweights* are constantly being updated. If i and j re-encounter after a short period of time, they can share the same previous information therefore avoiding to disclose more historical information. Ideally, upon an encounter between nodes i and j, the anonymized neighboring graph of j should only contain information of common nodes it has with i. This information is useful for i to update its ego network.

Binary anonymization consists in replacing the *pweight* associated to a given link with 1 or 0, if the weight is above or below a given anonymization threshold (ρ), respectively. This technique converts the weighted (randomized or not) neighborhood graph into an unweighted one, therefore hiding the *pweight* associated to a given edge. The selection of ρ is also limited by the utility of the neighboring graph. Consider, for example, that node a has nodes b, c and d as its neighbors with *pweights* ($w_{\mathrm{a,b}} = 0.05, w_{a,\mathrm{c}} = 0.15, w_{a,\mathrm{d}} = 0.65$). If ρ is set to 0.1, the anonymized *pweights* are ($w^*_{a,b} = 0, w^*_{a,c} = 1, w^*_{a,d} = 1$). But, if instead ρ is set to 0.25, the anonymized *pweights* would become ($w^*_{a,b} = 0, w^*_{a,c} = 0, w^*_{a,d} = 1$). If node a meets another node, say node e, a would tell e that its neighbors are ($w_{a,\mathrm{c}} = 1, w_{a,\mathrm{d}} = 1$) for $\rho = 0.1$ and ($w_{a,\mathrm{d}} = 1$) for $\rho = 0.25$.

C. Determination of routing metrics

Previous work [2] succeeded in identifying social structures, but the routing performance is affected as they did not take into consideration the dynamics of the network, i.e., the making and breaking of social ties. If, for instance, social similarity is considered, it is important to map actual interactions among nodes into social connectivity graphs comprising only stable social contacts in order to improve forwarding performance.

PRIVO represents the dynamics of the social structure as time-varying weighted neighboring graphs, where the weights (i.e., social strengths among nodes) express the average separation period over different timeslots.

1) Ego betweenness centrality

In PRIVO, each node's ego network corresponds to its neighboring graph if the *pweigths* are above a given *weight threshold* (ε). Since the connections among the ego direct neighbors are also necessary for the ego network, each node shares its anonymized neighboring graph (as explained in Section IV.B) with its neighbors.

Given a set of configuration parameters (see Section V for more details), the determination of ε can be seen as an optimization problem consisting in finding the ε that maximizes (or

minimizes) a certain routing performance metric (e.g., finding ε that maximizes the delivery ratio).

2) Weighted similarity to the destination

Let $\boldsymbol{\mathcal{A}}_n$ be the weighted adjacency matrix of node n at a given timeslot. Let $\boldsymbol{\mathcal{A}}_{n_{i,j}} = w_{i,j}$. If nodes i and j have met before, then $w_{i,j} \neq 0$; otherwise, $w_{i,j} = 0$. The weighted similarity of n to a destination node d (s_d) is obtained by summing the non-zero row entries in $\boldsymbol{\mathcal{A}}_{n_{i,d}} | i \neq n$. If n never met d but node i belonging to n's neighboring graph did, n may infer that i is a more suitable forwarder to d than him through i's anonymized neighboring graph.

3) Mean time to encounter

Besides *pweight*, PRIVO also uses a metric called *mean time to encounter* (MTTE) to determine the best message forwarder to a given destination taking into account the average separation period at each timeslot and the expected time necessary for the two nodes to re-encounter. Specifically, given that in PRIVO each node keeps an estimate of the average separation period at each timeslot that is updated as nodes encounter each other, PRIVO predicts the most probable timeslot for future contacts also taking into account the shortest time to re-encounter. As an example, consider that node a meets nodes b and c at 2pm and 5pm for 10 and 15 minutes, respectively. At 8pm, node a receives a message destined to node d that is expected to meet nodes b and c on the next day. When node a computes the average separation periods of b and c, it also considers the time to re-encounter nodes b and c in the following day assuming that these nodes maintain similar habits.

D. Routing algorithm

This section describes PRIVO's routing algorithm, i.e., the messages exchanged using the Paillier homomorphic encryption scheme and the routing decision process.

1) The attribute privacy mechanism

PRIVO ensures attribute privacy, as regardless of the metric ($\mathfrak{m}$) used by the routing algorithm (*pweight*, similarity to the destination, or ego betweenness centrality – $\{w_{i,j}, s_d, c_{EBC}\} \subset \mathfrak{m}$), when two nodes meet they find the best forwarder in a private manner using the Paillier homomorphic encryption scheme.

Let A be a node carrying a set of messages $\mathcal{M}$ and node B be a neighbor of A. Let $A \rightarrow B : < message >$ denote a message sent from A to B. Upon an encounter, A wants to know if B is the best forwarder to carry $m \in \mathcal{M}$ destined to D. Let p_k and s_k be public and private key, respectively.

The exchange of messages in PRIVO works as follows:

0. Node A calculates metric $\mathfrak{m}$ for each $m \in \mathcal{M}$ using the information it has available.
1. Each time A establishes a contact with another node, it announces: $-\mathfrak{m}_i \forall D_i \subset m_i | \; i = 1,2, \ldots |\mathcal{M}|$, the destination of the message and its public key p_{k_A} to B. Node A multiplies the metric $\mathfrak{m}_i$ by -1 to reduce the number of cryptographic operations to be performed by node B.

$$A \rightarrow B : < \mathcal{E}_{p_{k_A}}(-\mathfrak{m}_{A_i}), D_{m_i}, p_{k_A} >$$

2. Node B performs for each metric received the following operations: first, B sums $-\mathfrak{m}_{A_i}$ to the corresponding metric $\mathfrak{m}_{B_i}$, then it multiplies the result by a random one-use number ($nonce$) to randomize it. Without the multiplication, A would be able to obtain $\mathfrak{m}_{B_i}$ (by summing $\mathfrak{m}_{A_i}$ to the received non-randomized result). Then B sends the result $\mathcal{R}_i = nonce \cdot (-\mathfrak{m}_{A_i} + \mathfrak{m}_{B_i})$ to A.

$$B \rightarrow A : < \mathcal{E}_{p_{k_A}}(\mathcal{R}_i) >$$

3. A decrypts the received comparisons for each m_i.

$$\mathcal{D}_{s_{k_A}}\left(\mathcal{E}_{p_{k_A}}(\mathcal{R}_i)\right)$$

Node A knows that if $\mathcal{R}_i > 0 \rightarrow \mathfrak{m}_A < \mathfrak{m}_B$ which means that node B is the best forwarder. If that is the case, A forwards m_i to B.

$$A \rightarrow B :< m_i >$$

Obtaining the best forwarder can be demanding in terms of CPU, energy, etc., due to the number of messages that have to be exchanged in the process. In PRIVO, each node has a secure forwarding table (SFT) containing entries $< DestinationNode\ (DN), BestForwarder\ (BF) >$ that is updated each time a node meets another one that is a better forwarder than him. When the average separation period between two nodes is updated, if one of those nodes is a BF in the SFT, the entry is removed. SFT allows to reduce the number of messages exchanged when two nodes meet therefore reducing also PRIVO's consumption of resources.

2) The routing decision process

Because of the different routing metrics considered here, four variants of PRIVO are proposed. PrivoASP uses as routing metric *pweight*. PrivoMTTE uses as routing metric the mean time to encounter. PrivoSDBC, which is the social version of PRIVO, uses as routing metric weighted similarity to the destination and ego betweenness centrality. PrivoCOMBINED is a combination of PrivoMTTE and PrivoSDBC. It results from multiplying the routing metrics of PrivoMTTE and PrivoSDBC.

However, independently of the routing metric used, all buffered messages to be forwarded whose next forwarders are in the SFT are sorted based on their TTL, i.e., priority is given to new messages.

PrivoSDBC first compares the nodes' weighted similarity to the destination, in a secure manner. Only if the previous are equal is the ego betweenness centrality considered. Therefore, the message is sent first to the most similar node to the destination of the message and then to the more central node, if both nodes have the same similarity.

V. Simulation Model

PRIVO was implemented in the Opportunistic Network Environment (ONE) simulator [21]. Different simulation scenarios consisting of two synthetic mobility models and two real mobility traces were considered. It is assumed here, as in most networks of interest, that there is some social structure between

the nodes participating in the network. Each source node generated a new message according to the following intervals: 0.5 to 1min (0.5-1), 1 to 2min (1-2), 2 to 4min (2-4), 4 to 8min (4-8), 6 to 12min (6-12) and 8 to 16min (8-16). The length of the timeslots varied from 5, 10, 15, 30 and 60 min corresponding to 288, 144, 96, 48 and 24 timeslots per day, respectively. Similarly to values normally used in the estimation of the Round Trip Time (RTT) on the Transport Control Protocol (TCP) [22], α and β are set to 0.125 and 0.25, respectively.

A. *Synthetic mobility models*

The simulation time was 7 days with an update interval of 1.0 s. A map-based mobility model of the Helsinki city over an area of 4.5 × 3.4 Km was used. The message size varies from 500 kB to 1 MB. Only two nodes within range can communicate with each other at a time. The communication range between nodes was 10 m, and the communication was bidirectional at a constant transmission rate, for Bluetooth and Wi-Fi interfaces, of 2 Mbit/s and 10 Mbit/s, respectively. From time to time, a source node randomly chosen generated one message to a randomly chosen destination. Two mobility modes were considered:

1) Shortest-path Map-Based Movement (SPMBM)

SPMBM consisted of a network with 40 pedestrians, 20 cars and 6 trams. Pedestrians were moving at a speed varying between 0.8 to 1.4 m/s. Cars and trams were moving at a speed varying between 2.7 to 13.9 m/s. Each time a tram reaches its destination, it paused for 10 to 30 s. The TTL attribute of each message was 5 h. The pedestrians and cars had a buffer size of 10 MB. Trams had a buffer size of 100 MB for DTN traffic.

2) Working Day Movement (WDM)

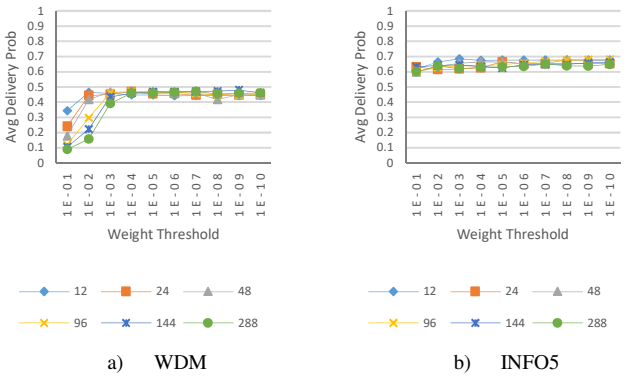
WDM consisted of a network with 100 pedestrians and 18 buses. There were 50 offices and the working day length was 8 h. The probability of going shopping after work was 50% and there were 10 meeting points. Pedestrians and buses were moving at a speed varying between 0.8 to 1.4 m/s and 7 to 10 m/s, respectively. Each time a bus reaches its destination, it paused for 10 to 30 s. The TTL attribute of each message was 24 h. All nodes had a buffer size of 20 MB for DTN traffic.

B. *Real mobility traces*

The haggle-one-infocom2005 (INFO5) [23] and taxicabs in Rome (TR) [24] traces, across different network and mobile environments, are used to provide additional support to the analysis and findings of this paper. In INFO5, 41 iMotes were distributed to students attending Infocom 2005 over 2.97 days. TR contains GPS coordinates of approximately 320 taxis collected over 30 days of taxicabs in Rome, Italy. The simulation duration and number of nodes of TR were reduced to 3 days and 304 nodes, respectively. All nodes had a buffer size of 10 MB for DTN traffic. The TTL attribute of each message was 24 h.

VI. Simulation Results

In this section, several simulation results describing the performance of PRIVO are presented. For each setting, i.e., protocol-configuration parameter pair, fifteen independent simulations using different message generation seeds were conducted,



a) WDM b) INFO5

Fig. 1 Delivery ratio of PrivoSDBC for WDM and INFO5 scenarios with η varying from 24 to 288 and ε varying from 1×10^{-1} to 1×10^{-10}.

and the results averaged, for statistical confidence. PRIVO was compared with well-known DTN routing protocols [11]: two non-social-based routing protocols, namely Epidemic [25] and Prophet [26], and two social-based routing protocols, namely BubbleRap [27] and dLife [28].

The four variants of PRIVO were considered: PrivoASP, PrivoMTTE, PrivoSDBC and PrivoCOMBINED.

The performance of PRIVO was evaluated according to the following metrics: delivery ratio, overhead ratio and cryptographic cost. The delivery ratio is a key performance indicator as it tells the percentage of successfully received packets of all sent. The overhead ratio is the number of message transmissions for each created message. The cryptographic cost, because of homomorphic encryption, gives the computation and transmission cost incurred by cryptographic operations.

In addition, information loss (or data utility) due to the use anonymization methods will also be evaluated. This will be accomplished by analyzing the correlation coefficients between a non-anonymized version of PRIVO and the anonymized ones over the simulations.

A. *The selection of the parameters: number of timeslots (η) and weight threshold (ε)*

This section analyses the selection of two important configuration parameters η and ε. In Fig. 1, the influence of η and ε in PrivoSDBC is analyzed through simulation for a synthetic (WDM) and a real (INFO5) scenario. In these scenarios, source nodes were generating messages every 6 to 12 min, ε varied from 1×10^{-1} to 1×10^{-10} and η varied from 24 to 288.

The delivery ratio in all scenarios increased with the reduction of ε and η. In general, it starts low as many links are ignored because of ε's high value and as ε reduces it increases and tends to stabilize, starting to decrease again as ε becomes very small (i.e., $\varepsilon \to 0$).

The highest delivery ratio was obtained when η was 144 and 96 for WDM and INFO5, respectively. In terms of ε, 1×10^{-7} and 1×10^{-10} provided the highest delivery ratio for WDM and INFO5, respectively. However, the gains for the same η and different ε were below 3% from 1×10^{-7} to 1×10^{-8} for WDM and 0.4% from 1×10^{-8} to 1×10^{-9} for INFO5, if compared to the highest ones.

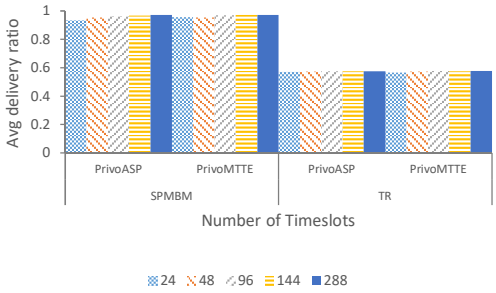


Fig. 2 Delivery ratio for PrivoASP and PrivoMTTE in SPMBM and TR scenarios with η varying from 24 to 288.

The weight threshold can be dynamically adjusted in a similar manner to the TCP congestion window [22] but starting at a high value, e.g., 1×10^{-3}, and gradually reducing it in a time-interval basis. If at the end of a given time interval the delivery ratio increased, then ε is reduced and vice-versa.

Now, the goal is to make a comparative analysis of PrivoASP and PrivoMTTE in terms of η since they do not depend on ε. Fig. 2 presents the delivery ratio for PrivoASP, PrivoMTTE in different scenarios.

Generally, two tendencies can be observed from Fig. 2 depending on the metric used. On the one hand, if estimates of the average separation periods are considered, the increase of η results in a slight increase of the delivery ratio. This is a direct result of having timeslots of smaller length, which offer estimates that are more accurate. PrivoMTTE uses these estimates and its performance slightly increases with the increase of η in all scenarios. As previously stated, MTTE allows to identify among all existing timeslots the best ones, that is, the ones with the highest value of average separation period and smallest duration to the next re-encounter, assuming, for example, that nodes' movements obey a certain pattern.

On the other hand, the other PRIVO variants use *pweight*, i.e., an average of the estimates of the average separation period. In this case, two behaviors were observed in Fig. 2 . The best performance was with $\eta = 288$ and $\eta = 96$ for SPMBM and TR scenarios, respectively. This was influenced by the contact patterns, which were more frequent in the WDM scenario.

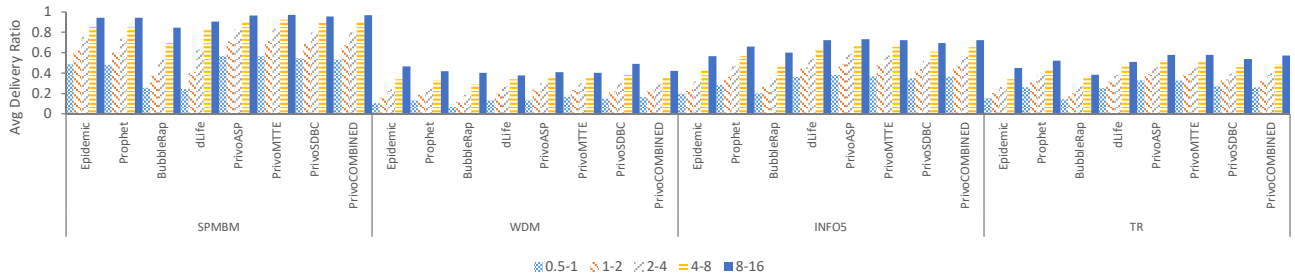
The increase of η also leads to disadvantages as more slots require more storage. Nevertheless, it is possible in all the PRIVO variants except PrivoMTTE to reduce storage by only keeping an estimate of *pweight* that is updated at the end of each timeslot, therefore not being necessary to keep estimates of the average separation period for each timeslot.

B. *Routing performance*

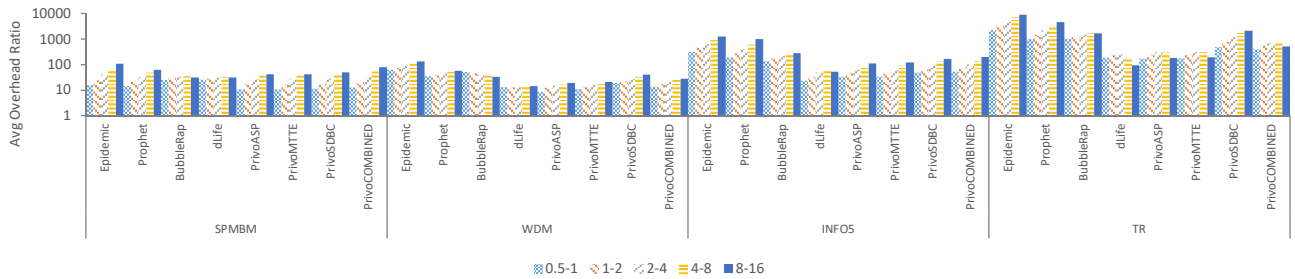
This section analyses PRIVO's routing performance without the use of homomorphic encryption. Based on the previous section, η and ε were set, respectively, to 144 and 1×10^{-8} for all PRIVO variants.

Fig. 3 presents the average delivery ratio and overhead ratio for different scenarios, routing protocols and message generation rates. As expected, with the decrease of the data rate there is an increase in delivery ratio and a decrease of the overhead ratio as fewer messages circulated in the network.

Overall, PRIVO performed better than other routing protocols in all message generation rates and scenarios in terms of delivery and overhead ratios. However, the performance of



a) Delivery ratio



b) Overhead ratio

Fig. 3 Delivery and overhead ratios for all the routing protocols considered in different scenarios for different message generation rates.

TABLE 1
AVERAGE PAILLIER EXECUTION TIMES (MS)

Key Size	$\mathcal{E}(a)$	$\mathcal{D}(c)$	$\mathcal{E}(a+b)$	$\mathcal{E}(a-b)$	$\mathcal{E}(k \cdot a)$
512	1.73 ± 0.0342	1.74 ± 0.0314	0.01 ± 0.0005	0.38 ± 0.02	0.02 ± 0.0016
1024	11.03 ± 0.1261	11.29 ± 0.3552	0.03 ± 0.0019	0.74 ± 0.0425	0.05 ± 0.0023
2048	83.49 ± 0.3029	83.9 ± 0.4546	0.06 ± 0.0033	1.74 ± 0.0719	0.14 ± 0.0038

TABLE 2
AVERAGE DELIVERY RATIO LOSSES AND GAINS USING THE PAILLIER CRYPTOSYSTEM (%)

	512	1024	2048
SPMBM	-0.08	-0.01	1.01
WDM	1.11	4.77	21.73
INFO5	0.00	-0.21	-0.09
TR	-0.11	-0.06	-0.88

TABLE 3
AVERAGE DELIVERY RATIO LOSSES AND GAINS USING THE PAILLIER CRYPTOSYSTEM WITH 1024 BITS KEY (%)

	SPMBM	WDM	INFO5	TR
PrivoASP	-0.04	4.61	-0.06	-0.09
PrivoMTTE	0.34	4.05	0.04	0.14
PrivoSDBC	-0.17	5.72	-0.50	-0.35
PrivoCOMBINED	-0.15	4.69	-0.32	0.06

each PRIVO variant depends on the scenario. The routing protocols that presented the highest delivery ratio were PrivoMTTE for SPMBM and TR, PrivoASP for INFO5 and PrivoSDBC for WDM. The maximum gains obtained were 14.6%, 29.9%, 29.8% and 49.5% for SPMBM, WDM, INFO5 and TR, respectively. Among the non-PRIVO routing protocols, the ones that presented the highest delivery ratios were Epidemic for SPMBM and WDM, dLife for INFO5 and Prophet for TR. Therefore, if there are some repetitive movement patterns then PrivoSDBC is the best choice otherwise, it is PrivoMTTE.

C. Cryptographic costs

This section analyses the cryptographic cost of using the Paillier homomorphic encryption scheme.

1) Additive homomorphic encryption

A set of experiments were performed to evaluate the performance of additive homomorphic encryption using the Paillier cryptosystem. The experiments were performed in a personal computer with the following specifications: Intel® CORE™ i7-2600 CPU @ 3.40GHz, 16 GB RAM and Windows 10 Pro (64-bits). Table 1 presents the average Paillier execution time of five operations, namely encryption $\mathcal{E}(a)$, decryption $\mathcal{D}(c)$, sum $\mathcal{E}(a+b)$, difference $\mathcal{E}(a-b)$ and multiplication by a constant $\mathcal{E}(k \cdot a)$. The difference is performed by multiplying the second term by -1 followed by summing the numbers, therefore being slower than sum and multiplication by a constant. The operations were repeated 100 times.

2) PRIVO's performance with Paillier

Now, messages were generated every 6 to 12 min. Table 2 presents PRIVO's average delivery ratio losses (+) and gains (-) using the Paillier cryptosystem with key sizes of 512, 1024 and 2048 bits for $\eta = 144$. Each table entry results from averaging losses and gains of all PRIVO variants per key. From Table 2, it is possible to see that the losses are below or equal to 1% in all scenarios, with exception of WDM, therefore the use of the Paillier homomorphic encryption was not considered in the previous subsection (Section VI.B).

A more detailed analysis was performed for the legacy key size (i.e., 1024 bits) [29]. Table 3 presents the average delivery ratio losses (+) and gains (-) using the Paillier cryptosystem for $\eta = 144$.

It was concluded, based on simulation results, that if a message was not transmitted because of the additional delay caused by homomorphic encryption, it would be transmitted later on. In some cases (see Table 2 and Table 3), this additional delay is beneficial to the routing protocol, as it may contribute to the reduction of the network load, even though the maximum achieved gains being negligible (at most 0.50% for the legacy key).

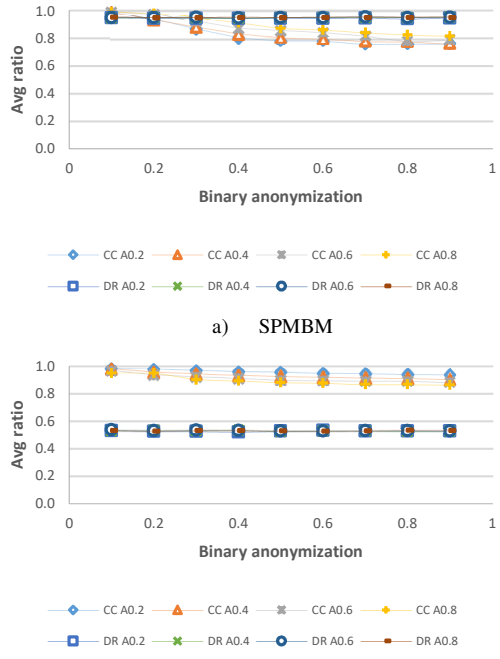
D. Information loss

This section analyses the utility of the data (or information loss) because of the use of anonymization methods. Information loss is measured comparing the correlation coefficients [30] of the ego betweenness centrality values of all the nodes in the simulation with and without anonymization. The ego betweenness values were collected at the end of each day and the values were compared for different percentages of total anonymization with the case were no anonymization was used. Total anonymization corresponds to the total number of nodes in the neighboring graph that are anonymized. Binary anonymization was applied over a percentage of the latter. At the end of each simulation, the correlation coefficients were averaged taking into account the number of days of the simulation. Different percentages of binary and total anonymization were used. The former varied from 10% to 90% with increments of 10% and the latter varied from 20% to 80% with increments of 20%.

Fig. 4 presents the average correlation coefficient and delivery ratio (DR) for PrivoSDBC in SPMBM and TR scenarios. Between binary anonymization and neighborhood randomization, the former is the one to cause a reduction on the average correlation coefficients as it increases, and this effect worsens as the percentage of total anonymization increases. Nonetheless, since PrivoSDBC uses ego betweenness centrality and weighted similarity to the destination and the latter is more frequently used as a routing metric, the effects of the lowest values of correlation coefficients (i.e., 0.82 for SPMBM and 0.86 for TR corresponding to 90% of binary anonymization and 80% of total anonymization) are not significant as can be seen by the steady average delivery ratio in Fig. 4.

VII. CONCLUSION

This paper proposed PRIVO, a PRIvacy-preserVing Opportunistic routing protocol for DTNs. PRIVO ensures link privacy



a) SPMBM

b) TR

Fig. 4 Average correlation coefficient (CC) and delivery ratio (DR) for PrivoSDBC in SPMBM and TR scenarios for different percentages of total (A) and binary anonymizations.

by means of binary anonymization and neighborhood randomization, and attribute privacy by means of the Paillier homomorphic encryption scheme.

The effectiveness of PRIVO is supported through extensive simulations with synthetic mobility models and real mobility traces. Simulations results show that PRIVO presents on average cryptographic costs below 1% in most scenarios, and if there are some repetitive movement patterns then PrivoSDBC is the best choice, otherwise it is PrivoMTTE. Furthermore, PRIVO presents on average gains of 22.2% and 39.7% in terms of delivery ratio for the synthetic and real scenarios considered, respectively.

A comparative analysis of PRIVO with other privacy-preserving schemes was left for future work. A threat model to evaluated PRIVO's resilience against link disclosure attacks, by eavesdropping the exchange of anonymized neighboring graphs between two nodes to disclose neighboring information, and attribute disclosure, by trying different anonymization thresholds to obtain the real *pweights*, was left for future work. In addition, an analysis of PRIVO's cryptographic costs with stronger keys sizes such as 3072 and 4096 bits was left for future work.

REFERENCES

[1] M. M. J. M. Khabbaz, C. M. C. Assi, and W. F. W. Fawaz, "Disruption-Tolerant Networking: A Comprehensive Survey on Recent Developments and Persisting Challenges," *Commun. Surv. Tutorials, IEEE*, vol. 14, no. 2, pp. 607–640, Jan. 2012.

[2] K. Wei, X. Liang, and K. Xu, "A survey of social-aware routing protocols in delay tolerant networks: applications, taxonomy and design-related issues," *Commun. Surv. Tutorials, IEEE*, vol. 16, no. 1, pp. 556–578, Jan. 2014.

[3] N. Magaia, A. Francisco, P. Pereira, and M. Correia, "Betweenness centrality in delay tolerant networks: a survey," *Ad Hoc Networks*, 2015.

[4] N. Magaia, P. Rogerio Pereira, and M. P. Correia, "Security in Delay-Tolerant Mobile Cyber Physical Applications," in *Cyber-Physical Systems: From Theory to Practice*, D. B. Rawat, J. J. P. C. Rodrigues, and I. Stojmenovic, Eds. CRC Press, 2015, pp. 373–394.

[5] X. Wu, X. Ying, K. Liu, and L. Chen, "a Survey of Algorithms for Privacy-Preservation of Graphs and Social Networks," *Manag. Min. Graph Data*, p. 37, 2009.

[6] X. Lu, P. Hui, D. Towsley, J. Pu, and Z. Xiong, "Anti-localization anonymous routing for Delay Tolerant Network," *Comput. Networks*, vol. 54, no. 11, pp. 1899–1910, 2010.

[7] A. Shikfa, M. Önen, and R. Molva, "Privacy and confidentiality in context-based and epidemic forwarding," *Comput. Commun.*, vol. 33, no. 13, pp. 1493–1504, Aug. 2010.

[8] G. Costantino, F. Martinelli, and P. Santi, "Privacy-preserving interest-casting in opportunistic networks," in *IEEE Wireless Communications and Networking Conference, WCNC*, 2012, pp. 2829–2834.

[9] M. Radenkovic and I. Vaghi, "Adaptive user anonymity for mobile opportunistic networks," in *7th ACM International Workshop on Challenged Networks, CHANTS 2012*, 2012, pp. 79–81.

[10] I. Parris and T. Henderson, "Privacy-enhanced social-network routing," *Comput. Commun.*, vol. 35, no. 1, pp. 62–74, 2012.

[11] A. Sánchez-Carmona, S. Robles, and C. Borrego, "PrivHab+: A secure geographic routing protocol for DTN," *Comput. Commun.*, vol. 78, pp. 56–73, 2016.

[12] R. L. Rivest, L. Adleman, and M. L. Dertouzos., "On data banks and privacy homomorphims," *Found. Secur.*, 1978.

[13] N. Zeilemaker, Z. Erkin, P. Palmieri, and J. Pouwelse, "Building a privacy-preserving semantic overlay for Peer-to-Peer networks," in *Proceedings of the 2013 IEEE International Workshop on Information Forensics and Security, WIFS 2013*, 2013, pp. 79–84.

[14] P. Paillier, "Public-Key Cryptosystems Based on Composite Degree Residuosity Classes," *Adv. Cryptol. — EUROCRYPT '99*, pp. 223–238, 1999.

[15] M. Everett and S. P. Borgatti, "Ego network betweenness," *Soc. Networks*, vol. 27, no. 1, pp. 31–38, Jan. 2005.

[16] E. M. Daly and M. Haahr, "Social network analysis for routing in disconnected delay-tolerant MANETs," *Proc. 8th ACM Int. Symp. Mob. ad hoc Netw. Comput. - MobiHoc '07*, p. 32, 2007.

[17] M. Mcpherson, L. Smith-Lovin, and J. M. Cook, "Birds of a Feather: Homophily in Social Networks," *Annu. Rev. Sociol.*, vol. 27, no. 1, pp. 415–444, 2001.

[18] M. J. Freedman, K. Nissim, and B. Pinkas, "Efficient Private Matching and Set Intersection," *Eurocrypto 2004*, vol. 3027, no. i, pp. 1–19, 2004.

[19] G. Zhong, I. Goldberg, and U. Hengartner, "Louis, Lester and Pierre: Three Protocols for Location Privacy," *Pets'07*, pp. 62–76, 2007.

[20] T. Hossmann, T. Spyropoulos, and F. Legendre, "Know Thy Neighbor: Towards Optimal Mapping of Contacts to Social Graphs for DTN Routing," in *2010 Proceedings IEEE INFOCOM*, 2010, pp. 1–9.

[21] A. Keränen, J. Ott, and T. Kärkkäinen, "The ONE simulator for DTN protocol evaluation," in *Proceedings of the Second International ICST Conference on Simulation Tools and Techniques*, 2009, p. 55.

[22] J. Kurose and K. Ross, *Computer Networking - A top-down approach*. Pearson, 2012.

[23] D.-G. Akestoridis, "{CRAWDAD} dataset uoi/haggle (v. 2016-08-28): derived from cambridge/haggle (v. 2009-05-29)." 2016.

[24] L. Bracciale, M. Bonola, P. Loreti, G. Bianchi, R. Amici, and A. Rabuffi, "CRAWDAD dataset roma/taxi (v.2014-07-17)," *CRAWDAD wireless network data archive*. 2014.

[25] A. Vahdat and D. Becker, "Epidemic routing for partially connected ad hoc networks," Technical Report CS-200006, Duke University, 2000.

[26] A. Lindgren, A. Doria, and O. Schelén, "Probabilistic Routing in Intermittently Connected Networks," Springer Berlin Heidelberg, 2004, pp. 239–254.

[27] P. Hui, J. Crowcroft, and E. Yoneki, "BUBBLE Rap: Social-Based Forwarding in Delay-Tolerant Networks," *IEEE Trans. Mob. Comput.*, vol. 10, no. 11, pp. 1576–1589, Nov. 2011.

[28] W. Moreira, P. Mendes, and S. Sargento, "Opportunistic routing based on daily routines," in *2012 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, WoWMoM 2012 - Digital Proceedings*, 2012, pp. 1–6.

[29] ENISA, "Algorithms, key size and parameters report 2014," 2014.

[30] R. A. Fisher, "Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population," *Biometrika*, vol. 10, no. 4, pp. 507–521, May 1915.