This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI

10.1109/ACCESS.2020.3009691, IEEE Access

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

**A Survey on Privacy Issues in Mobile Social Networks**

**AKSHITA MARADAPU VERA VENKATA SAI1, (Student Member, IEEE), YINGSHU LI1,(Senior Member, IEEE)**

* Department of Computer Science, Georgia State University, Atlanta, GA 30303, USA

Corresponding authors: Akshita Maradapu Vera Venkata Sai (amaradapuveravenkat1@student.gsu.edu) and Yingshu Li (yili@gsu.edu).

This work is partly supported by the National Science Foundation (NSF) under Grant Nos. 1741277 and 1829674.

 **ABSTRACT** Mobile Social Networks (MSNs) are a category of social networks that have features of bothtraditional OSNs and Location-Based Social Networks (LBSNs). MSNs provide users with an opportunity to explore social media, connect with people nearby, or with those sharing similar interests. The ease with which people can connect and use these networks to obtain information, share user-centric data, and request services on the go through their mobile devices as long as they have access to the internet has made MSNs omnipresent. Due to the widespread nature of MSNs and the lack of effective privacy-preserving architecture makes them a target for several attacks by adversaries and the users of the networks are completely vulnerable to these privacy breaches. With the advancements in technology, the number of attacks and their possible solutions is complex and expansive. Therefore, there is a need to properly categorize them to understand the intricacies of different attacks and the concepts used by researchers to devise different solutions. This survey on privacy issues in MSNs serves as a way to provide an introduction to MSNs. We categorize privacy in MSN on the basis of the aspect they effect and summarize the different threats in each category. Furthermore, we provide an elaborate classification of the various privacy-preserving solutions proposed in recent research works, under each category of privacy. We also review different datasets and data generation tools that can be used by future researchers in this field.

 **INDEX TERMS** Differential Privacy, location privacy, Mobile Social Networks, privacy, user privacy.

**I. INTRODUCTION**

HE past two decades have seen an exponential increase Tin the production and use of mobile devices like mobile phones, tablets, and the establishment of several new stan-dards in the supporting mobile communication technologies like cellular data networks, Bluetooth and Wifi. Mobile com-munication was previously restricted to text messaging with a character-limit per message and voice calls made over the cellular data networks. With time and advancements in the field of mobile communication, the cellular service providers have started creating data plans for the internet which allowed people to connect and surf the web on-the-go.

The Internet started gaining popularity in the 1990s and became a more accessible commodity to the general public. Easy availability meant that people started using it more often to explore content on the “Web”. During this time, people identified a new use case for the internet, which was to explore their neighborhood to identify other people sharing similar interests, with an intention to expand their social

network and reach. In light of these events, the first social networking site called Geocities [1] was created in the year 1994. This social network allowed users to create personal websites which were then grouped by cities based on the content posted on the website.

Social networking sites became more popular and gained attention with the advent of MySpace [3] in 2003 followed by the creation of LinkedIn and Facebook in 2004 [2] [4]. These different social networking sites became part of a much larger network called Online Social Networks (OSNs). The users of OSNs could connect to the network by logging into their account through a browser-based web application. Unfortunately, these web applications could not be used effectively on mobile devices as it was hard to adapt them to smaller screens on these devices due to the lack of supporting mobile UI technology. This encouraged developers to design tools and create mobile versions of these social networks. These networking sites and several other service applications that run on mobile devices then came to be known as Mobile

|  |  |
| --- | --- |
| VOLUME x, 20xx | 1 |

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3009691, IEEE Access



Author et al.: Preparation of Papers for IEEE TRANSACTIONS and JOURNALS

Social Networks (MSN).

MSNs inherit a lot of features from Delay Tolerant Net-works (DTNs) and Opportunistic Networks (OppNets) re-garding prolonged connectivity and exploiting social charac-teristics to make new connections. MSNs can be described as a type of DTN where users make connections with each other and share content with common interests. When it comes to how MSNs are similar to other social networks, they inherit the social features of traditional OSNs, and neighborhood and location features from Location-Based Social Networks (LBSNs). MSNs today mainly constitute of the following kinds of applications:

Messaging applications: These applications allow users to send messages and share content like images and documents with people in their contact list or emails. These applications are centered around conver-sations. Whatsapp, WeChat, Facebook Messenger are examples of these applications.

Neighborhood discovery applications and Location Based Services (LBS) : These applications exploit the user’s current location or location history to provide location-based updates or neighborhood information like restaurants, malls, theatres and many more. For example, depending on the places one visits frequently, like Home and Work, Google Maps regularly sends up-dates on the best way to reach the place. It also provides on-the-go traffic data, highlighting denser traffic routes, alternate routes and the places the person has already visited in the neighborhood. Fandango, for instance, gives information on the movies being played in the theaters close by.

MSNs require abundant user information like name, age, email id, hobbies, Point of Interest (POI) to improve service quality, and this information is requested periodically to stay up-to-date. As MSNs have both social and mobile aspects, it is extremely difficult to come up with a scheme that provides privacy protection and retains service quality. The lack of a privacy-preserving architecture allows adversaries to access sensitive information making users vulnerable to privacy breaches. Following are a few potential privacy threats to MSNs:

1. Identity theft: This includes exploiting a user’s per-sonal information to create impersonation profiles on social networks or to steal monetary possessions. One such incident was the data breach occurred in Equifax in 2016 [5], that led to a leak of user’s sensitive credit card information like their social security number, name, credit information and much more. This data breach led to a lot of counterfeited transactions that brought upon losses to both users and the company. In [22], Sweeney, showed how a released anonymized social data can still be analyzed to trace back to a specific individual.
2. Location tracking: With a lot of location-based appli-cations that store user location information and share

them publicly, outsiders can obtain information about a user’s daily activities and trajectories easily. This can then be further analyzed to profile users and launch many other attacks.

1. Fake profiles: This is a very common occurrence when it comes to platforms like Facebook or Instagram where a lot of fake profiles are created with the inten-tion to contact people and lure them into giving away sensitive information. These profiles sometimes leave back viruses and trojans on the devices that are used to open messages from the said profile.
2. Malicious links: These links are usually shared through messages and emails, which leads the recip-ients to phishing websites and trick them into giving away sensitive information like passwords or credit card information, or automatically download a pretend software onto the device which might launch certain background attacks. Most times the users are com-pletely unaware of these attacks.
3. Hacking: This is the most common and popular issue in social networks, where adversaries, also known as “Black Hats” can make their way into a user’s account and procure complete control over it.

The remainder of this paper is organized as follows. Section

1. covers MSNs’ components and architecture. Section III formulates MSN as a graph and its social features. The different categories of privacy in MSNs and the potential threats in these categories are covered in Sections IV and V respectively, while the proposed solutions and their clas-sification are provided in Section VI. Section VII lists and summarizes the available real-world datasets and synthetic dataset generators. Section VIII and Section IX conclude the paper and discuss the future research directions in the field, respectively.
   1. **MSN COMPONENTS AND ARCHITECTURE**

**A. COMPONENTS IN MSN**

A MSN is a dynamic environment, therefore, its compo-nents are constantly going through changes in states and the connections they form. Thus, an MSN at any given point can have several types of components like users, groups and communities, different mobile infrastructures like Bluetooth, cellular data, or the basic internet. The main components of an MSN can be narrowed down to the following:

1. Nodes or Mobile Devices: These include all devices of mobile nature that the end-users might use to connect to an MSN. It can be a mobile phone, a tablet, laptop, or any device that can establish a Bluetooth connection and is embedded with sensors.
2. Network: This includes the infrastructure or the frame-work that is used to establish connections and serve as a communication medium among end-users and between end-users and servers. The infrastructure might include servers, routers, access points, cellular base stations, and also cloud infrastructure if the services being pro-vided are hosted on the cloud.

2 VOLUME x, 20xx

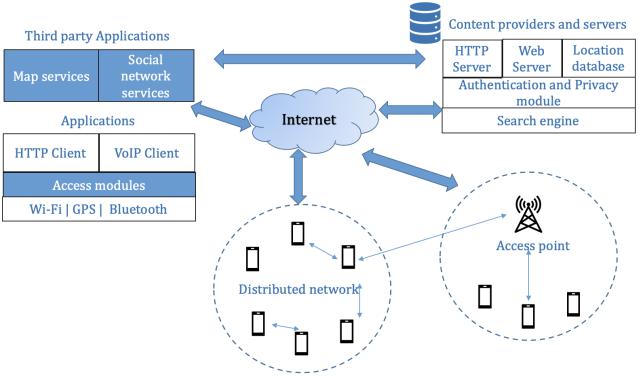
This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3009691, IEEE Access



Author et al.: Preparation of Papers for IEEE TRANSACTIONS and JOURNALS

|  |  |  |
| --- | --- | --- |
| obtained as follows, |  |  |
| 1; | if wi;j > 0 |  |
| Ai;j = (0; | otherwise | (2) |



We will now define and formulate the social features of an MSN.

**A. NODE DEGREE**

The degree of a node tells us the number of connections the current node (user) has with other nodes (users) within the graph G (network). The degree of a node i can be obtained as follows,

**FIGURE 1.** MSN Architecture

1. Content Providers and Servers: Content providers in-clude all the servers that the users may directly or indirectly send requests to and also other third-party servers that make content available to end-users.

k

|  |  |  |
| --- | --- | --- |
| Xj | (3) |  |
| di =Ai;j |  |
| =1 |  |  |

**B. NODE STRENGTH**

Node Strength gives information about the frequency of communications between a node (user) and the other nodes (users) in the graph G (network). The node strength of a node i can be obtained as follows,

**B. ARCHITECTURE**

Depending on the way the users communicate with each other and with the server, MSNs can either have a central-ized, decentralized, or hybrid architecture. Most MSNs today follow a hybrid architecture as shown in Fig. 1, because they are either mobile versions of older web applications or they want to exploit the benefits of the hybrid architecture. The centralized aspect of hybrid architecture provides benefits like simple implementation, reduced hardware, low main-tenance costs, and high efficiency while the decentralized or distributed aspect provides benefits of low reliability and stability requirements, better traffic load balancing, lower latencies and presence of alternate sources of data in case of a server failure.

**III. MSN AS A GRAPH AND SOCIAL FEATURES**

In order to resolve the different privacy issues, one should clearly understand the structure of MSN and it’s social as-pects from a technical stance. This can be done by formu-lating MSN as a graph. For simplicity let’s assume that the graph is undirected, and is denoted as G = (K; W ). The users of an MSN are denoted by nodes K = f1,2,: : :, i,: : :, kg. The edges represents connections between two nodes and the weight of each edge can be obtained as follows,

|  |  |  |  |
| --- | --- | --- | --- |
| wi;j = | Fi;j | (1) |  |
| T |  |
|  |  |  |

where Fi;j is the frequency of communications between two nodes i and j within a given time period T . In order to derive the other features, we need an adjacency matrix. This adjacency matrix will help us describe the nature interactions among users in the network. As we have an undirected graph, the adjacency matrix A, will be symmetric and it can be

k

|  |  |  |
| --- | --- | --- |
| Xj | (4) |  |
| si =Ai;jwi;j |  |
| =1 |  |  |

**C. CLUSTERING COEFFICIENT**

This feature of the graph sheds light on how tightly social groups and communities are connected. Higher the clustering coefficient, greater the aggregation relationship between a node and its surrounding nodes. This feature further helps us define social groups and virtual communities. If di is the node degree of node i then the number of edges (Ni), between node i and the other nodes is given by,

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Ni = | di(di 1) | | (5) |  |
| 2 |  |  |
|  |  |  |  |

The local clustering coefficient of node i can be obtained as follows,

|  |  |  |
| --- | --- | --- |
| k | (6) |  |
| Ci = jP |  |
|  | wi;j |  |
| =1 |  |  |

Ni

The global clustering coefficient can be calculated as an average of local clustering coefficients of all nodes and can be obtained as follows,

|  |  |  |
| --- | --- | --- |
| k | (7) |  |
| C = iP |  |
|  | Ci |  |
| =1 |  |  |

jKj

**D. BETWEENESS CENTRALITY**

In a network, there is always exchange of information be-tween a source S and destination D. In the absence of a direct path between S and D, the information passes through some node i. A few nodes always act as intermediate nodes and few would never be an intermediate node. This feature helps us compare the importance of a few nodes over the

|  |  |
| --- | --- |
| VOLUME x, 20xx | 3 |

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3009691, IEEE Access



Author et al.: Preparation of Papers for IEEE TRANSACTIONS and JOURNALS

others in an MSN. A greater betweeness centrality implies that more available paths between any source and destination are passing through node i. The betweeness centrality of a node i can be obtained as follows,

|  |  |  |  |
| --- | --- | --- | --- |
| bi = | i | (8) |  |
| 2 S;D |  |

where i is the number of paths that pass through i and S;D is the number of paths between source S and destination D.

**E. CLOSENESS CENTRALITY**

This feature indicates the amount of interaction of a node with the other connected nodes in the network. A greater value indicates higher number of interactions and more fre-quent communications among the nodes. For node i, the closeness centrality can be calculated as follows,

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cc(i) = |  | k 1 | (9) |  |
|  |  |  |
|  | k |  |
| P | |  |  |
| j=1 d(i; j) |  |  |

where d(i; j) is the length of the path between node i and node j.

**IV. TYPES OF PRIVACY IN MSNS**

Privacy in MSNs can be broadly classified into three cate-gories:

1. Location Privacy: User’s location is extremely impor-tant for many services provided by an MSN. Preserv-ing user’s location information without affecting the quality of service is crucial, and these location privacy-preserving solutions are highly dependent on the ap-plication. A few applications like share rides require a single location while other applications like navigation systems, require continuous location inputs. Solutions for location privacy-preservation include perturbation [70], obfuscation [71], k-anonymity [49], spatial cloak-ing [72] and temporal cloaking [73] to name a few.
2. User privacy: Users in an MSN can be considered as nodes of an MSN graph. Therefore, user information includes both node information and the information regarding the links they establish within the network. The goal of user privacy-preservation solutions is to preserve both node and link information connected to the user [51]. Node and edge perturbation [74], and anonymization [75] are a few common methods used in designing user privacy-preserving solutions [45].
3. Communication privacy: This mainly focuses on pre-serving the content and the context of the information being exchanged in a network or uploaded to the server and other third-party applications. This information may include message contents, a few attributes of user’s profile information, queries made to the server that may include sensitive information like userID or user location, and many more. Most communica-tion privacy-preserving mechanisms use session key agreements and digital signatures [65], authentication and verification schemes [64], and different encryption methods [23], [76].

**V. KNOWN THREATS TO MSNS**

Advancements in MSN technology have made more re-sources available to adversaries to design attacks that pose threats to different aspects of an MSN. It is vital to understand the nature of these attacks and their underlying mechanism to design privacy-preserving solutions. In this section, we categorize the threats to location, user, and communication privacy in MSNs.

1. **THREATS TO LOCATION PRIVACY**
   1. Direct sharing attacks: These attacks happen when users share their location directly on social platforms like Facebook and Foursquare as check-ins or Geo-tagging. On these platforms, attackers have direct ac-cess to user’s check-in history, therefore, they need not build a model to extract users’ location information. For this reason, these attacks are passive in nature. Due to the availability of original location data, the attack models, in these cases are easy to build and are very effective. [?] [31] describe extracting location information and launching location-based attacks.
   2. Continuous tracking: This is a more active approach compared to direct sharing attacks. These attacks are designed to geo-locate a user without any prior infor-mation and they are based on indirectly shared loca-tions like the obfuscated location. Sharing obfuscated locations is very popular in Facebook Marketplace and Wechat [31]. As these are popular applications, the amount of obfuscation used is publicly available. This allows attackers to create a model to reverse engineer the exact location and study the user’s move-ment patterns. These are known as spatial knowledge attacks [34]. Another type of continuous tracking hap-pens in neighborhood-discovery services, where the users might want to explore places or events in the neighborhood. Here, the user’s location is sent to the LBS (Location based server) as part of a search query. This location can then be inferred either by direct query sampling [35] or by sending multiple queries to identify the exact location of the user [33].
   3. Inference attacks: Inference attacks are defined as the analysis of data to gain knowledge of the subject un-lawfully. These attacks are more common as most mo-bile applications have location sharing capabilities and do not impose strong privacy mechanisms to protect this information [53]. In [36], the authors were able to perform an inference attack on GPS data gathered from their volunteers. They were able to exactly identify and locate the homes of these volunteers by analyzing the GPS data and segmenting it into discrete trips. They then implemented four heuristic algorithms to identify the exact location of the subject’s home.
   4. Spoofing: One of the most common types of spoofing is geo-location spoofing [11], where a user fakes his location with malicious intent and is primarily done by either using VPNs or a DNS Proxy. GPS spoofing is

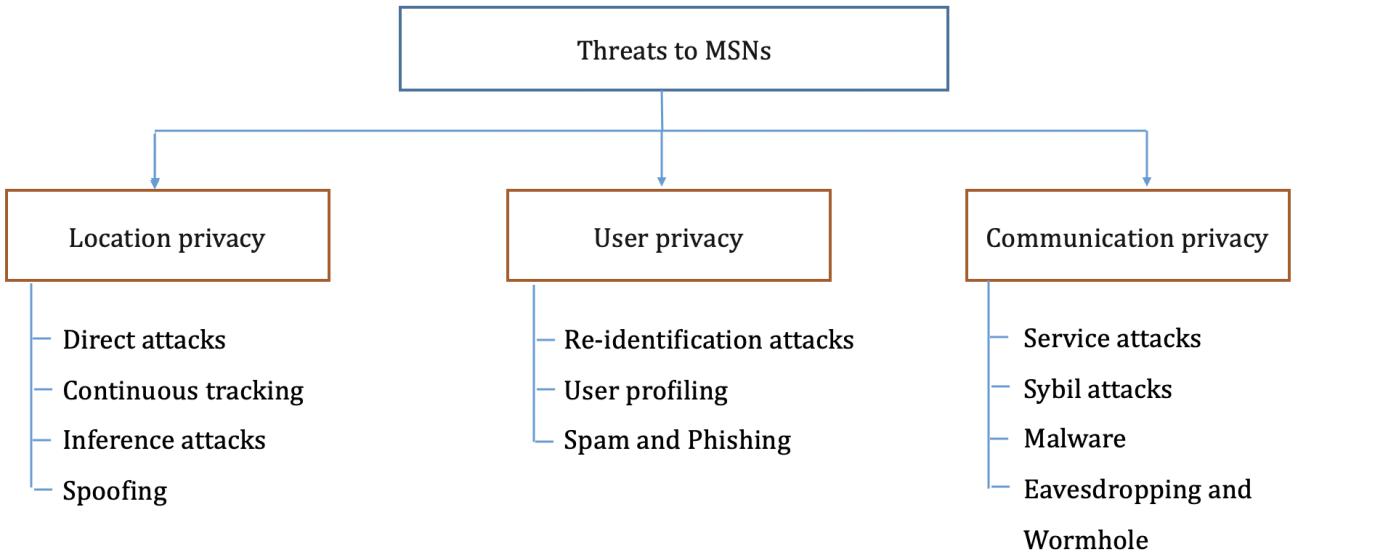
4 VOLUME x, 20xx

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3009691, IEEE Access



Author et al.: Preparation of Papers for IEEE TRANSACTIONS and JOURNALS



**FIGURE 2.** Threats to privacy in MSNs

another type of spoofing that aims at deceiving users by broadcasting incorrect GPS signals, delaying GPS signals, or re-sending these signals in a different zone.

1. **THREATS TO USER PRIVACY**
   1. Re-identification attacks: These attacks use multiple data sources and common “quasi-identifiers” in these sources to uniquely identify a user [54]. Several studies show how public anonymized medical data can be used to uniquely identify patients [77]. For example, in [22], the author was able to uniquely identify about 87% of users by combining the medical and voter registration data and just three identifiers. Similarly, authors in [38] were able to identify users from four spatio-temporal points. In [39], the authors designed a re-identification attack which uses points of a user’s mobility trace obtained from a trace dataset like Ge-oLife and Cabspotting, to form a heat-map structure. From the GeoLife trace set, they were able to uniquely identify about 80% of the users.
   2. User profiling: In MSNs, users are either matched to other users or events based on the attributes they share, like user location or interests. This matching process sometimes requires users to publish their information, which may include sensitive data. The attackers can also study user behaviors online, from their check-in data, posts, and many more [52]. All the leaked information combined with the availability of powerful data mining and analysis techniques allows attackers to profile users or gather their information [43]. The most common way of misusing user information is the cre-ation of fake profiles, impersonation and identity theft, which by extension can contribute towards launching targeted attacks [10].
   3. Spam and Phishing attacks: Spam and phishing at-tacks work together and are a common way of obtain-ing user’s confidential information. These attacks work by sending legitimate looking emails or messages that lead users to a website with some embedded logic to extract all the information entered. While spam directs users to a third-party website, phishing attacks directs users to an almost perfect looking website. These at-tacks have become increasingly common in MSNs like WhatsApp, LinkedIn, Facebook, Twitter, and many more. On platforms like Facebook and Twitter, at-tackers impersonate famous brands, celebrities, and sometimes customer support. All of these pages lead users to a legitimate looking official website and either ask users to answer a few questions or to enter their information in a form on the website [17]. Another famous spam was on WhatsApp, where a message was circulated about ‘offering a free pair of Adidas trainers to celebrate their 93rd anniversary’. Similar to the previous case, the message had a link that directed the users to legitimate looking Adidas webpage where they were asked to enter their credit card information [18] [19].
2. **THREATS TO COMMUNICATION PRIVACY**
   1. Eavesdropping and wormholes: Eavesdropping is when an attacker intercepts the communication be-tween two parties without their knowledge. In wired communications this attack happens over the network but, when it comes to MSNs to launch this attack, a malicious code is embedded into the application. There are several social networks and other applica-tions on App Store and Google Play like Facebook that is programmed to listen to audio through the

|  |  |
| --- | --- |
| VOLUME x, 20xx | 5 |

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3009691, IEEE Access



Author et al.: Preparation of Papers for IEEE TRANSACTIONS and JOURNALS

mobile device’s microphone. These apps listen to the audio from television shows or ads and mine the audio data to send targeted advertisements to the users [12]

* 1. Sometimes, social networks like Facebook also listen to users’ conversations as part of social media monitoring and send targeted advertisements to them

A wormhole attack, on the other hand, is where an attacker gets packets from one point, tunnels them to another point and either replays them later or directs them to a different location. These attacks make use of users’ network IDs and can further lead other attacks like Man-in-the-middle (MITM), where the recorded packet information may be changed and forwarded or to replay attacks. These attacks are designed to disrupt routing protocols or other security protocols being used in MSNs [37]

1. Service attacks: These attacks aim at making a re-source unavailable or disrupting the service. Denial of Service (DoS) and Replay Attacks are examples of such attacks. DoS or DDoS (Distributed Denial of Service) is a cyber-attack that infects multiple devices by injecting malware to attack servers. Several MSN applications like Facebook, LinkedIn, Uber, Airbnb, banking applications, and e-commerce applications are vulnerable to this attack due to the ease with which an attacker can profile a user. A Dos attack on a mobile phone happens through an application that is down-loaded onto the device. This application either directly performs a DDoS attack or opens up a security loop-hole in the way that the attacker has complete control over the device. This attack primarily reduces revenue to companies by blocking network traffic and incurs additional costs to them to overcome the issue. WireX botnet [15] is one such application that was dubbed as an "Android Clicker" and, affected over 120,000 Android devices and conducted massive DDoS attacks in the application layer. Another application, Mirai botnet affected a lot of social networks like Amazon and Twitter [16]. A replay attack is a network attack where valid communication information is collected and then replayed or delayed. It is a version of the Man-in-the-middle attack.
2. Malware: In these attacks, users are directed to a website that either automatically downloads malicious code or requests users to download a supporting media player or an application or enable access to cookies to continue viewing the page. This in fact is some malicious code that when downloaded and/or installed allows the attackers to control mouse and keyboard activities of the infected device. The distribution of malware in MSNs happens through fake profiles [20] or broadcasted messages that show up directly in the user’s inbox [21].
3. Sybil Attacks: In this type of attack, a malicious user also known as Sybil may create several fake

identities to gather information about users and attack the communication network itself. These attacks are particularly prevalent in MSNs due to its open and distributed architecture. Sybils are then used to launch phishing and DoS attacks to distribute malware. One of the major attacks by sybils is on routing protocols, where the sybils place themselves in a way, such that several individual paths between different sources and destination pass through them [41]. In another attack, Sybils establish connections with other Sybils and honest nodes and then start disseminating spam, advertisements, and malware to violate user privacy. Additionally, sybils can generate different reviews to favor their services or undermine other services, and this is done by focusing on some specific behaviors and repeating them at high frequency [40] [42].

**VI. PROPOSRED SOLUTIONS**

Extensive research has been conducted in recent years to address and resolve privacy issues in MSNs. These solutions have a few similarities with respect to concept, technique or the feature of MSN being preserved. To carry out further research in this field, it is imperative to have a clear under-standing of the work done so far. Therefore, in this section, we provide an elaborate classification of privacy-preserving solutions as shown in Fig. 3 and summarize them.

**A. LOCATION PRIVACY**

Based on the technique used in location privacy-preserving mechanism, the solutions can be categorized into : 1) K-anonymity based schemes; 2) Obfuscation based schemes; and 3) Differential privacy based schemes

1) K-anonymity based schemes

K-anonymity is a popular technique that is used in several privacy solutions and was first proposed in [22]. According to [22], a scheme is k-anonymous if the probability of uniquely identifying a particular entity from k entities is at most 1=k. Several solutions in this area that use k-anonymity have been proposed and they either considered Online Social Networks (OSNs) or involved trusted third parties (TTP). Also, most MSNs store the user information and location information on separate servers, which then requires some sort of encryption to safely link the two servers and provide accurate query results. This type of distributed architecture increases the risk of information leakage.

In [48] authors proposed a mechanism called CenLoc-Share to address the above mentioned issues. Firstly, to overcome the issues caused by having two different servers was solved by combining the Social Network Server (SNS), and Location Based Server (LBS), into a single server called Location Storing Social Network Server (LSSNS). Then, the authors identify scenarios where the user might send his location data to LSSNS For each scenario, the user submits a query to LSSNS, during which he sends his location along with (k-1) dummy locations, thus using k-anonymity. The

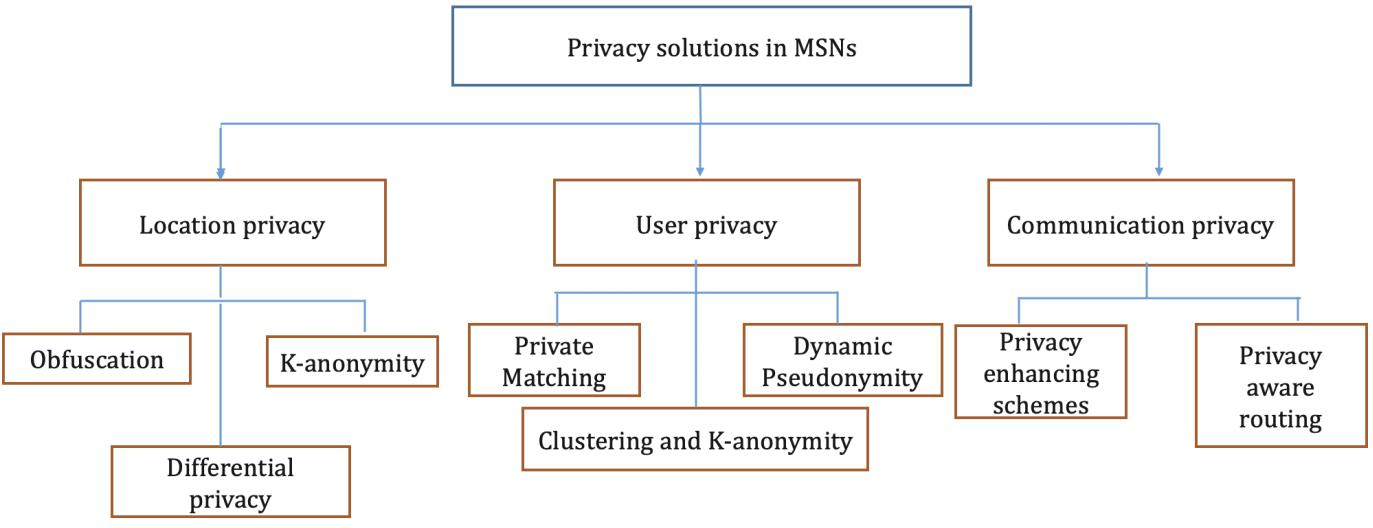
6 VOLUME x, 20xx

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3009691, IEEE Access



Author et al.: Preparation of Papers for IEEE TRANSACTIONS and JOURNALS



**FIGURE 3.** Privacy Solutions in MSNs

main contributions of this work are; it provides a central-ized scheme by using a single server, which reduces the risk of information leakage, it designed a scenario-specific LBS query processing instead of a generalized solution, uses “Sequence ID" in queries to prevent replay and tampering attacks and finally, reduces the storage requirement and the time needed to process queries compared to other mecha-nisms. Although the scheme provides several benefits, the centralized approach increases the computation complexity as the network increases. Also, the scheme only preserves the privacy of a single location. If the user submits continuous queries it might be relatively easy for an attacker to map the trajectory of the user, as the radius within which the dummy locations are generated is fixed.

As mentioned in the previous solution, queries sent to the server may increase the risk of information leakage. Therefore, in situations where the user sends continuous queries to the server, there is an increased risk to the location information. To address this, a solution called Collabora-tive trajectory privacy-preserving scheme was proposed in [50]. The basic idea of the scheme is to preserve location privacy by reducing the number of queries being made to the LBS by exploiting the caching ability of the user devices in the network. The scheme contains two algorithms :

1. Multi-hop caching aware cloaking: This allows the user to communicate within a Hmax (maximum hop distance) by sending a collaboration request. The users who respond to this request share their cached infor-mation. Based on this information, users can create a k-anonymous cloaking region and also locally obtain query results for what should be the next location, instead of sending a query to the LBS. The scheme provides different versions of this algorithm for the requestor and receiver of the “collaboration request".
2. Collaborative privacy-preserving querying: This al-

lows users to obtain information locally from the data cached and shared by other users, or it can send a query to a remote LSP. In order to obtain accurate results, the algorithm checks for the freshness of cached data. It also allows users to send fake queries to the LBS with the farthest location in its cloaking region, to introduce confusion.

The scheme can be used in both static and continuous querying scenarios. The k-anonymity in creating the cloaking regions and the confusion introduced by the fake queries pro-vide two-fold privacy preservation. But, the major disadvan-tage of this scheme is that the algorithms are computationally intensive, and may not be feasible for a mobile device in MSNs due to their limited computation capacities.

In [49], a solution similar to [50] called Privacy Preserving System (PPS) was proposed. This solution overcomes the computation challenges faced in [50]. To minimize queries being made to the LBS, PPS is used to maintain a single cache instead of multiple caches, with all the frequent loca-tion requests and their corresponding results. When a user makes an LBS request, the PPS checks if the location in the query meets either of the following conditions and then returns the result from the cache :

A request is being made within a small acceptable distance or,

A request is being made from a place that is a subarea of already cached locations.

If none of the above conditions are met by the location, the PPS directs the request to LSP by obfuscating the user’s location. The obfuscation radius is selected in a way that the region is k-anonymous. If the region has sparse users, then dummy users are added to make it k-anonymous. Compared to all the other solutions, this scheme, firstly, provides privacy to continuous queries, thus preserving the trajectory of the user. Secondly, it overcomes the computation issues pre-

|  |  |
| --- | --- |
| VOLUME x, 20xx | 7 |

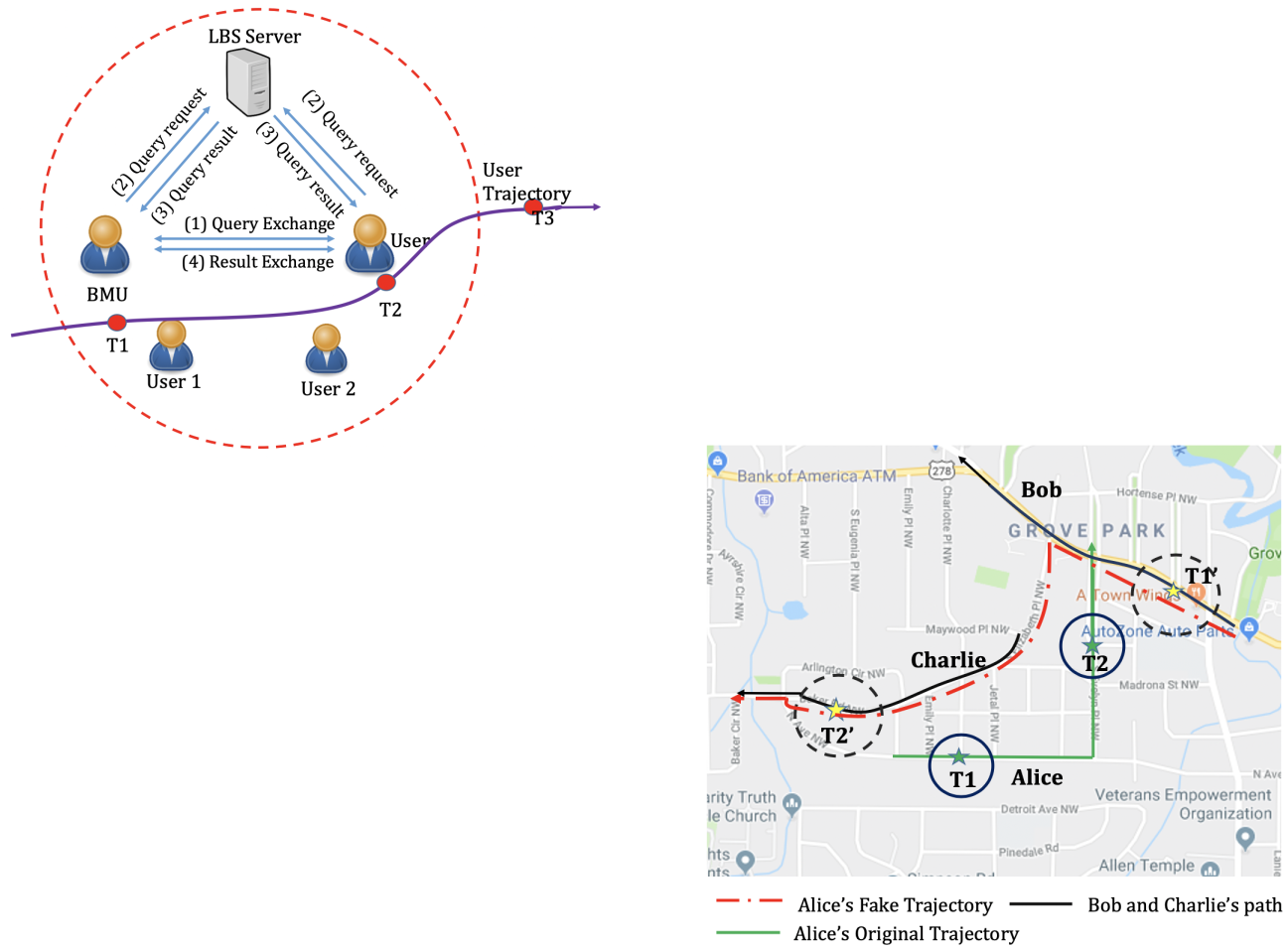
This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3009691, IEEE Access



Author et al.: Preparation of Papers for IEEE TRANSACTIONS and JOURNALS

is difficult to link a user to the ID because the BMUs keep changing as the user’s moves. Thirdly, the solution provides an additional level of privacy by obfuscating the location before submitting the query to the LSP. Finally, the query results are encrypted using asymmetric encryption making it resistant to eavesdropping attacks. Fig.5 shows how DQE preserves a user’s location trajectory. But, the authors do not consider the varying speeds of the users, making it difficult to apply this solution to a more practical scenario, and they do not discuss how often the scheme finds BMUs. Finally, as the solution is carried out at the user level, these calculations and multiple communication exchanges may drain the limited power capacity of the mobile devices.



**FIGURE 4.** DQE Mechanism

sented by the previous two schemes, by using a single cache to store all the results. This not only reduces the amount of data that needs protection but also reduces the number of communication exchanges and requests between users, resulting in less power consumption. Finally, the solution provided is more practical as it considers real-world user distributions, and how it is not always possible to have enough users to provide k-anonymity and proposes a way to overcome this obstacle as well.

In the recent years, the number of users using MSNs has increased, therefore, when solutions based on k-anonymity are implemented in practical scenarios, if the cloaking region is small, then the tendency of disclosing location informa-tion is more. Therefore, we need solutions based on other techniques like obfuscation which are not subject to user distribution.

2) Obfuscation based schemes:

In [46], the authors proposed a privacy preserving scheme called Deviation based Query Exchange (DQE) to preserve the user’s trajectory data. The scheme preserves privacy at the user level and has the following steps.

1. Step 1: Finding the Best Matching User (BMU).
2. Step 2: Deviation based Query Exchange (DQE).

In Step 1 user U is matched with other users and the best match possible is identified. During this step, to ensure user privacy, private matching [Section VI-B2 ], is performed by introducing confusion to the user’s original information (x; y; movement direction). A similarity value is calculated to see how similar a user is to others, and the user who is the least similar to U is selected as the BMU. In Step 2 shown in Fig. 4, U exchanges his ID with the BMU. The BMU will then forward U’s location query to the LSP by obfuscating the location and the obtained results are exchanged later. This solution is extremely well rounded and thought through because, firstly, the obfuscation happens at the user level, in-stead of at a central point, thus avoiding having a single point of failure/attack. Secondly, even if an attacker gets the ID, it

**FIGURE 5.** BMU Selection in DQE

To overcome the drawbacks of the DQE scheme, [44] pro-posed a model called SmartMASK which machine learning to build a fine-grained location privacy system. The model works as follows :

1. A clustering algorithm is used to generate the user’s location profile by using the user’s mobility history and location profiles.
2. Each check-in in mobility history has a different sensi-tivity level. Based on this, a trained Classification and Regression Tree (CART) model assigns a privacy level ( low, medium or high) to the check-in.
3. The users choose their location sharing preference (coarse or fine-grained location).
4. Based obfuscation level and the user’s preference, the obfuscation engine performs a hybrid obfuscation technique that includes the obfuscation operators: ra-dius enlargement, radius reduction, and center shifting. When the predicted privacy level is "high", a simple cloaking is applied along with hybrid obfuscation.

Unlike the previous solution, SmartMASK is centralized and thus, more capable of handling intense computations due to more available resources. Only privacy-level prediction and application of the obfuscation level is done for each new

8 VOLUME x, 20xx

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3009691, IEEE Access



Author et al.: Preparation of Papers for IEEE TRANSACTIONS and JOURNALS

location which is much less intensive. But, as the model performs hybrid obfuscation on the location, the utility of location data may be decreased considerably.

To reduce the utility loss from obfuscation, authors in [47] propose an ML-based model to learn the user motivation behind a location check-in. The proposed method, firstly, takes the location check-ins and already available user moti-vation to train a model that predicts future motivation. Then, users then provide information on the effect of different obfuscation level on their check-in utility. Based on these re-sponses and the predicted motivation labels, a cost-sensitive decision tree model (J48) is trained to predict the user’s perceived privacy level. This solution is the first of its kind as it considers user-specific utility while designing the model that does not use differential privacy. Firstly, it addresses the effect of obfuscation on utility and specifically trains models to predict a privacy level that retains the highest amount of utility. Secondly, designing such intelligent models relieves users from making sensitive and critical privacy decisions. The major drawback here would be the unavailability of these types datasets for future researchers.

3) Differential Privacy based schemes:

Differential privacy (DP) has become the gold standard in privacy. Unlike most other privacy-preserving solutions, dif-ferential privacy based solutions’ main aim to retain data utility, they also assume that the attacker has complete knowl-edge of the users and the network, and finally these schemes can also quantify the level of privacy they provide. Differ-ential privacy based solutions can be successfully applied in places where aggregate information is published, and DP would require that the changes made one location of a user should have a negligible impact on the final output making it impossible to send useful information to the LBS.

To address this issue, Dewri proposed a method that uses both k-anonymity and differential privacy to preserve loca-tion information [55]. In this method, the author fixes an anonymity set consisting of k locations where the probability of reporting the same obfuscated location x from these k lo-cations is the same. To achieve this, the author adds Laplacian noise [8] to each Cartesian coordinate of the location. Though the choice of Laplacian noise is better to retain utility and has been extensively proved in the work, one of the major issues with this scheme is the selection of anonymity set that greatly affects privacy.

Another differential privacy based solution for preserving location privacy is LPT-DPk [56]. In this work, Yin et.al. focus on persevering frequent location patterns. The authors first create a frequent pattern tree called the Location Infor-mation tree based on the frequency of location check-ins. Once the tree is generated, the top-k frequent patterns are selected by using weighted selection based on Exponential mechanism [?] and a Laplacian noise [8] is added to the top-k frequent location patterns set to preserve privacy. The method is then evaluated against another top-k mechanism for the utility of the location data. The LPT-DPk Scheme retains

more utility of the data and has a relatively low and stable error compared to the other previously proposed DP based schemes, but, it does not discuss how the initial frequent patterns are derived, as this greatly impacts the effectiveness of the mechanism.

In [57], the authors propose a novel DP method that im-plements Reinforcement learning (RL) to preserve a node’s semantic trajectory. The scheme is designed using game model as well, with the nodes and adversary as players. RL is implemented to selected the optimal privacy budget for the DP scheme. The obtained optimal budget is used to generate the “gamma noise" which will then be added to the location. This obfuscated location is then forwarded to the LBS and the results obtained as returned to the user. The implementation of RL in Location privacy is a relatively new concept and this paper effectively makes use of both RL and game model. Also, the optimal budget (strategy) is selected in a dynamic environment, unlike most other DP schemes, which on static location instances.

**B. USER PRIVACY**

User privacy aims at preserving a user’s profile information while communicating with other users or servers in an MSN. User privacy-preserving solutions can be broadly categorised into : 1)Clustering and K-anonymity ; 2) Private Matching ; and 3) Dynamic Pseudonymity.

1) Clustering and K-anonymity

In these schemes similar users are grouped together and privacy is provided at a group level to reduce information loss that might occur when techniques like Naive Anonymization are applied. SANGEERA [61] is one such clustering algo-rithm with the following procedure:

1. Nodes are partitioned into different clusters based on their quasi-identifiers and neighborhood information.
2. The quasi-identifier attributes are anonymized for each cluster to achieve k-anonymity.
3. All users in a cluster are collapsed to one node to prevent the revelation of intra-cluster nodes and edges.
4. Multiple relationships (edges) between two clusters are collapsed into one edge, to provide anonymity.

Previous anonymization solutions resulted in a significant information loss, whereas in SANGEERA, as edge gener-alization is adopted over perturbation, the structural infor-mation loss is reduced to a great extent. But, the major drawback of this scheme occurs when the clusters are very small or dense. In this situation even if the quasi-identifiers are anonymized the adversary can easily identify users based on other sources of user information. The algorithm does not consider the mobility and the dynamic nature of MSNs, where neighborhoods and communities change constantly making it inapplicable for more real-world MSNs.

To address drawbacks in solutions like SANGEERA, a new algorithm called Equicardinal Clustering was pro-posed in [62]. In this work, the user information is preserved

|  |  |
| --- | --- |
| VOLUME x, 20xx | 9 |

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3009691, IEEE Access



Author et al.: Preparation of Papers for IEEE TRANSACTIONS and JOURNALS

at the network level. The algorithm can be summarized as follows:

1. Similar users are clustered using k-means.
2. To achieve k-anonymity in each cluster, the users are reclustered as follows :
   1. Distance between the users and each cluster cen-troid is measured.
   2. Based on the distances, users are assigned new clusters.
   3. This is repeated until there are no more than n=k users are present in each cluster.
3. Users in a single cluster are represented by a cluster head.
4. All the links between two clusters are replaced by a single weighted edge. This weight represents the number of links between the two clusters.

This solution firstly reduces the information loss greatly compared to other schemes that use traditional clustering algorithms. As the neighborhood of the user is not considered for clustering, the privacy provided is not subject to user location, making the solution applicable to both OSNs and MSNs.

As most clustering algorithms used in these solutions are NP-hard, the solutions only provide sub-optimal results, therefore, cluster independent schemes like Private Matching need to be considered that utilize social features of an MSN to design a solution.

2) Private Matching

Friend Matching is a core feature of MSNs. MSNs allow users to connect with people in their neighborhood or net-work based on shared interests. To match users, they have to share sensitive information with the network which poses a serious threat to user privacy.

To solve the above issue a privacy-preserving profile matching scheme for MSNs called FindU was proposed in

1. In FindU three privacy levels are implemented for private matching between P1 and Pi, where 2 i N:
2. PL-1 : P1 and Pi will know the common attribute set.
3. PL-2 : P1 and Pi will only know the size m1;i of the the common attribute set.
4. PL-3 : P1 and Pi will only know the rank of each value

m1;i.

These privacy levels can be personalized by users, and the adversary can only obtain the output and the private inputs, thus, decreasing the amount of information he can obtain with each increasing level of privacy. In order to attain privacy levels, they designed two schemes :

1. Basic Scheme: This is defined to realize PL-1. In this scheme, a technique called Private Set Intersection (PSI) is used, where the attributes are encoded using hashing before the users share them.
2. Advanced Scheme: This is defined to realize PL-2 and PL-3. In this scheme, they use both PSI with BP (Blind-and-permute), where the user’s attributes are

encoded and the shared attribute sets are permuted so the link between the ranks and the attributes is broken. In this scheme, each sharing is then encoded using homomorphic encryption.

Once the protocol ends, P1, ranks m1;i locally to identify its best match, and then sends a connection request. This scheme takes into consideration every possible step at which the adversary may try to obtain information, and then designs schemes to make sure that the matching is resistant to active attacks. Also, the information of the user is being encrypted instead of being generalized, thus retaining the complete utility of the data. Though FindU provides several benefits compared to state-of-the-art privacy schemes in MSN like FNP and FC-10, it has higher communication costs as the encryptions are done on each communication between P1 and Pi at every step in the matching process.

In order to overcome the issues in solutions like FindU, a privacy-preserving profile matching mechanism called POSTER, was proposed in [59]. In POSTER, the secure matching is done by using perturbation.

1. All user attributes are converted to binary values to create profile vectors.
2. Mixed vectors for secure sharing are generated by adding noise to the profile vectors and performing a secure dot product of profile vectors of A and B who have to be matched.

After these initial steps are carried out, the authors created the following two schemes :

Basic Scheme: The secure dot product computations occur on the receiver end and in the presence of other users called Helpers. If the helper has both noise and the mixed vector, he might be able to obtain the user’s private information, thus this is collusion.

Collusion Resistant Scheme: To avoid collusions, the scheme divides noise or perturbation into chunks and send it to multiple helpers. They compute the dot product and sends it to B. B, then computes the final dot product to check its similarity with A.

This paper firstly does both secure friend matching as well as authentication by using a Verification scheme, to make sure that valid users are exchanging information. Also, as the mechanism does not use the computation heavy operations like Homomorphic encryption, it reduces the computation complexity and communication cost. The major drawback is that it does not focus on “helpers” selection which makes it vulnerable to Sybil attacks because when an adversary can act as multiple users, a few such profiles can be selected as helpers for the same communication allowing the adversary to obtain information even in the Collusion Resistant scheme of POSTER.

To avoid depending on other users, as in POSTER for private matching, Li et.al in [60] proposed a scheme called Match-MORE. This scheme was designed for users to se-curely match with friend-of-friends. The complete mecha-nism lies in the Matching degree function which uses Katz

10 VOLUME x, 20xx

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3009691, IEEE Access



Author et al.: Preparation of Papers for IEEE TRANSACTIONS and JOURNALS

Score and Dice similarity coefficient to calculate the social strength of two users, and the similarity score between two users, respectively. The matching happens in two phases :

1. Friend Discovery Phase: In this phase, A discovers new friends by broadcasting a connection request. Each responder sends their similarity score as a reply. A then selects the responder with the best score as the ‘friend’.
2. Friend Recommendation Phase: In this phase, the new friend goes through his friend list and calculates the similarity between A and all his friends, and then recommends a friend with the best similarity score to A.

In both phases, the similarity calculation is done using the Matching degree function and all the communications im-plement bloom filters. Unlike the schemes mentioned earlier, this paper quantifies the privacy provided by the scheme by using Shannon entropy and theoretically proves the accuracy, effectiveness, and efficiency of the scheme. Also, Match-More is lightweight as it completely avoids the use of Homo-morphic encryption. Finally, unlike other schemes that share actual attributes (true or perturbed) for matching, this scheme avoids that and just does matching using scores, therefore, reducing the risk of information leakage to the minimum.

3) Dynamic Pseudonymity:

Dynamic Pseudonymity Mechanism (DPP) proposed in

1. aims at providing both user and location information privacy. To protect the user’s identity, the scheme uses multi-ple anonymizers. The DPP scheme divides the user’s LBS queries into chunks, where each chunk is forwarded to a different anonymizer. While forwarding the query chunks belonging to the same user are assigned different pseudo-identities while interacting with different anonymizers. This makes sure that the adversary can not link user information to a query result or obtain a true User ID from any one of the anonymizers. The anonymizers also have a k-anonymity function to ensure user’s location privacy in the query. The paper discusses two threat models, “weak adversary" and “strong adversary" based on the locations of attack which are: 1) the wireless channel between the user and the LBS and 2) the anonymizer itself. The scheme was then de-signed to address these threats. The major advantages of this scheme are the fact the authors considered different levels of threats making the solution well-rounded. The use of hash trees and a single k-anonymity operation greatly reduces the computation time compared to other dynamic pseudonymity solutions. Also, the privacy that this scheme provides is two-fold privacy as it ensures both user and location privacy.

Another Dynamic Pseudonymity based scheme was pro-posed in [63] which focuses on understanding the context behind an LBS request to ensure user privacy. To impede the adversary from linking user ID to his true location, an identity management system is used. To make the system more secure, a hashing over pseudoID is performed, where the hash key is combined. The method proposed in paper [63]

enhances user privacy at two levels: 1) The identity manage-ment system replaces user identities with pseudo identities and this pseudoID is retained until the service request is fulfilled, thus for each service request, a new pseudoID is assigned to the user. 2) Hashing over user’s pseudo IDs by using userID + Service time as the hash key, to securely share the pseudo IDs. One of the main benefits of this scheme is that it has multiple levels at which privacy is ensured. Also, as the pseudo ID is hashed, this not provides user data privacy but also ensures communication privacy to a certain extent, as the pseudo ID is a part of the query packet being forwarded to the LBS. But, as the scheme generates new IDs and performs hashing for every query, and the computation is performed on the user-end, it might deplete the limited resources available on the mobile device.

**C. COMMUNICATION PRIVACY**

Depending on how a communication privacy-preserving mechanism is carried out, the solutions can be classified into : 1) Privacy enhancing schemes; and 2) Privacy-Aware Routing mechanisms

1) Privacy enhancing schemes

These schemes are designed to work alongside already im-plemented security protocols in MSNs and they heavily de-pend on Digital Signatures, Key Agreements, and certificates.

PRIF, proposed in [66] is an improved version of the forwarding scheme that is prevalent in MSNs today. The forwarding scheme proposed revolves around the concept of common interest-based communities formed in MSNs and, it preserves communication privacy by ensuring the privacy of the interacting parties. This is done by hiding the user’s interests and other information before he joins a community or interacts with anyone from the community. This privacy-preserving authentication protocol uses Schroff signatures and Group certificates handled by a central trusted authority TA. Though the scheme uses strong authentication protocols to ensure the privacy of the communicating entities, the use of central authority for the token generation is not ideal as it becomes a single point of failure for this scheme. Instead, if the scheme provides a way to improve trust between the communicating entities, it eliminates the need for central authorities and also adds a distributed aspect to the scheme.

Privacy Preserving Authentication Scheme (PPAS) [65] is a scheme based on group signatures. Group signatures are primarily used to provide user anonymity and un-linkability. This scheme ensures user legitimacy by generating an un-linkability token by using the group’s public key parameters. To ensure integrity and confidentiality during communica-tions the scheme uses signing and verification algorithms and session key agreement for every communication. This mechanism reduces the overload that the above-mentioned scheme suffers from, by reducing the number of tokens generated and considering group signatures instead. Unlike other schemes, it also considers the mobility of the user as part of the scheme, instead of a snapshot of the MSN. But,

|  |  |
| --- | --- |
| VOLUME x, 20xx | 11 |

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3009691, IEEE Access



Author et al.: Preparation of Papers for IEEE TRANSACTIONS and JOURNALS

similar to the previous scheme, this solution also depends on a single trusted authority to carry out the token generation.

As most privacy-enhancing schemes rely on trusted au-thorities to function which may be vulnerable to serious breaches, it is necessary to consider routing protocols that enhance privacy because they are more widespread in terms of the aspects they include and effect.

2) Privacy-Aware Routing mechanisms

textbfOnion routing [67] is one of the first strategies proposed to preserve communication privacy. This strategy ensures data integrity in both connection and connectionless systems. It uses mixers or onions, which store, encrypt, and forward data to the next node in random order. Generally, multi-ple mixers are used to ensure that communication is pro-tected against traffic analysis. Though this strategy ensures anonymous communication, the major drawback is the uni-directional feature of mixers or onions. This means that the mixers can only carry operations for one-way communica-tion. To make it bi-directional a set of reply onions needs to be deployed. The second drawback of onion routing is that it focuses on a single communication. An improvement over this method is proposed in [68]. This method extends onion routing to a multi-casting scenario and uses “Bloom filters” to enhance communication anonymity by obscuring the routing list of communication packets. It is one of the first works to use the concept of bloom filters in a privacy setup, and it also overcomes the disadvantages of the one-way privacy enhancement present in older solutions.

3PR [69], is a communication privacy-preserving scheme that uses machine learning techniques to learn user’s mobility patterns to predict their future routes and uses those predicted routes to route message. This is done by calculating the maximum likelihood of a node encountering the destination, and then, these likelihood values are hidden from other nodes within and outside the community. As part of the scheme, privacy-preserving functions like "max probability" and "par-tial sum" that make use of random number generators are pro-posed. This work uses the idea of “route-recommendations" in a communication setup which is novel and first of its kind. One major drawback of this solution is that preserves only the information related to the packet’s possible destination and fails to preserve the privacy of the packet’s content (message), which may hold sensitive information.

**VII. ONLINE RESOURCES AND DATABASES**

In this section, we will be listing and summarizing a few data sets and data generating tools available online that are used extensively in research on privacy in MSNs and LBSNs.

The most popular datasets are the Facebook [79] and Twitter [80] that available as part of SNAP (Stanford Net-work Analysis Project) by Stanford University [78]. The Facebook dataset [79] consists of users, their friend lists and ego networks. It was collected from the Facebook apps of survey participants and has 10 ego networks with a total of 4039 nodes and 88234 edges. Each user is represented by

**TABLE 1.** Summary of MSN and LBSN datasets

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Name | Domain | |  |  | Dataset information |  |
|  |  |  |  |  |  |  |
| Facebook | user | profiles | | and | Ego-networks:10 Nodes: 4,039 |  |
| Edges: 88,234 |  |
|  | friend lists | |  |  |  |  |
| Twitter | user | friend | | lists | Ego-networks:973 Nodes: 81,306 |  |
| Edges: 768,149 |  |
|  | and ego-networks | | | |  |  |
| Deezer | friend network | | | |  |  |
| Romania | with liked music | | | | Nodes: 41,773 Edges: 125,826 |  |
| Croatia | genres | |  |  | Nodes: 54,573 Edges: 498,202 |  |
| Hungary |  |  |  |  | Nodes: 47,538 Edges: 222,887 |  |
|  |  |  |  |  |  |  |
| Brightkite | check-ins | |  | and | Check-ins: 4,491,143 Nodes: |  |
| dataset |  | 58,228 Edges: 214,078 |  |
|  | friendship | |  |  |  |  |
|  | network | |  |  |  |  |
|  |  |  | |  |  |  |
| Gowalla | user | profiles, | | lo- | Check-ins: 36,001,959 Users: |  |
|  | cation profiles and | | | | 319,063 Locations: 2,844,076 |  |
|  | location check-ins | | | |  |  |
|  |  | | | |  |  |
| Weeplace | check-ins, profiles | | | | Check-ins: 7,658,368 Users: |  |
|  | and location infor- | | | | 15,799 Locations: 971,309 |  |
|  | mation | |  |  |  |  |
|  |  | |  | |  |  |
| Foursquare | location | | based | | Nodes: 106,218 Edges: 3,473,834 |  |
|  | friendship | |  |  |  |  |
|  | network | |  |  |  |  |
|  |  |  |  |  |  |  |
| GeoLife | user | GPS | trajec- | | Users: 182 Trajectories: 17,621 |  |
| Timespan: 3 years |  |
|  | tory |  |  |  |  |  |
| LifeMap | user location mon- | | | | Users: 8 Nodes: 9681 Edges: 1717 |  |
|  | itoring, user | | | tra- | Wi-Fi APIS: 52,510 |  |
|  | jectories | |  |  |  |  |
|  |  | | | |  |  |
| T-drive | taxi GPS traceset | | | | Users: 10,000 Timespan: 1 week |  |
|  |  | | | |  |  |
| Cabspotting | taxi GPS raceset | | | | Users: 500 Timespan: 30 days |  |
|  |  |  |  |  |  |  |
| Manhattan | taxi GPS traceset | | | | Trajectories: 1000 Timespan: 1 |  |
| Taxi | year |  |
|  |  |  |  |  |
|  |  |  |  |  |  |  |

25 features like location, education degree, job start date, and end date, employer information, workplace, and several others. These networks are represented by undirected graphs. The Twitter dataset [80] like the Facebook dataset, consists of friendship circles (lists) and ego networks, but the network here is represented as a directed graph with about 973 ego networks and a total of 81306 nodes and 768149 edges.

Another friendship dataset from SNAP [78] is Deezer [81], [82]. The dataset has friendship networks of users from 3 different European countries and contains three sub-networks represented by directed graphs for Romania, Croa-tia, and Hungary. The number of nodes and edges in each of these sub-networks are mentioned in Table 1. The dataset also provides user’s preferred genres preferred which have been compiled based on the songs users liked in a music network.

Brightkite [84] is an LBSN dataset that consists of user’s location check-ins and their friendship relationships. The data-set includes user check-ins and friendship networks of users within the social network. It has more than 4 million check-ins with 58,228 nodes (users) and 214,078 edges.

12 VOLUME x, 20xx

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3009691, IEEE Access



Author et al.: Preparation of Papers for IEEE TRANSACTIONS and JOURNALS

Each check-in includes user id, check-in time stamp, latitude and longitude readings, and location id. This dataset has been specifically used to study user friendships and mobility patterns in MSNs [83]. This particular dataset is sparser than other mobility datasets because we only have places at which users checked-in deliberately.

Gowalla [85] is a popular LBSN dataset collected using Gowalla API. For each user, the dataset has user profiles, their friendships, and location check-in history and each loca-tion an attached location profile. Based on these profiles, the locations are categorized into 7 subcategories like commu-nity, nightlife, entertainment and many more. Over the years, the data collected from the API comprises 36 million check-ins with 319,063 users and over 2,844,076 locations. This dataset has been mainly used in location recommendation systems [24], but can also be used to generate privacy models based on the location predictions.

The Foursquare dataset [86] is another LBSN dataset that is a part of the Social Computing Data Repository at Arizona State University and provides a friendship network among users of the Foursquare network. The social network was available to users with GPS enabled mobile devices and the data was collected through software installed on their respective devices. This dataset consists of 106,218 nodes and 3,473,834 edges.

Weeplaces [85] is another dataset obtained from a web-site. The website that was used to collect this dataset is now integrated into other social network applications like Facebook Places, Foursquare and Gowalla. The website was used to visualize user check-ins and the dataset generated includes user’s friends who use Weeplaces, location check-ins and additional information about the locations. This dataset contains 7,658,368 check-ins generated by 15,799 users over 971,309 locations. It is similar to the Brightkite and Gowalla datasets but provides additional information about the location like locationID, city, and location category.

GeoLife GPS Trajectories [25], [87] is a trajectory dataset collected by Microsoft Research Asia, Geolife project. It contains GPS trajectories of 182 users collected over three years (from April 2007 to August 2012). It has a total of 17,621 trajectories with location co-ordinates and the altitude of users’ locations and the locations were updated every 1 to 5 seconds. Datasets like these can be used to understand user’s mobility patterns [26] and design privacy models based on it. GeoLife has been used extensively to provide location privacy in Mobile Crowd Sensing (MCS) systems.

LifeMap Mobility data [6] is a dataset generated by a mo-bility monitoring system called LifeMap at Yonsei University in Seoul. The dataset contains fine-grained mobility data of 8 users collected over two months in Seoul, Korea. The dataset includes location coordinates, Wi-Fi fingerprints and user-defined places. The locations were collected by the system every 2 to 5 minutes. The data were collected from users’ mobile devices and has 9861 nodes on 1717 paths and 52510 Wi-Fi APIs. This dataset was used in research on mobility learning and movement predictions of an MSN user [27]–

[29].

T-drive [91], Cabspotting [90], and Manhattan Taxi tra-jectory [88] [89] are a few taxi trajectory datasets. Manhattan taxi trajectory has 1000 taxi trajectories collected over one year in the city of Manhattan [30]. Cabspotting is a traceset of mobility data of taxi cabs in San Francisco. It contains GPS coordinates of 500 taxi cabs collected over 30 days. T-Drive is a dataset collected as part of Microsoft Research, featuring taxi drivers in Beijing and has over 10,000 users with data collected over one week.

Apart from the datasets mentioned so far, there are also ways to generate synthetic data using online data generators. These tools provide users with several options and fields to generate custom datasets. Mackaroo [92] is one such website that lets us generate mock data with user profiles and a variety of other fields like location, occupation and so on. Generatedata [93] is another online data generator similar to Mackaroo. It lets us generate random users with a variety of fields along with their location data. The data can be generated in a plethora of formats like Excel, HTML, JSON, SQL, and XML. DTM Generator [94] is another popular data generator that produces data rows and schema objects and also other optional schema objects like views, triggers and many more. It is highly compatible with most popular database systems like MySQL, Oracle and Microsoft SQL server. Several other data generators have been mentioned and summarized in [94].

**VIII. CONCLUSION**

In this paper, we present a survey on privacy issues in MSNs. First, we introduced MSNs, and its components and architecture. Then, we formulated MSNs as a graph and, then defined and formulated its social features like Node degree and Centrality. We then categorized privacy in MSNs into three areas: location privacy, user privacy and communication privacy and listed the threats to each are. We then provided an elaborate classification and summary of the various solutions proposed by researchers to resolve the aforementioned privacy issues. As part of the study, we learned that the following things have to be considered while proposing a privacy-preserving scheme for an MSN :

The adversaries’ knowledge about the network and it’s users.

Computation requirements, as most of the mechanisms, are carried out on the user’s end device that has limited computation and storage capabilities.

Connectivity requirements, which implies that one needs to factor in the mobility of the users that may result in prolonged periods of disconnectivity.

Effect of the privacy-preserving mechanism on data utility.

Apart from the threats and solutions we also explored and listed several real-world datasets and data generators that can be used by future researchers. Most of these datasets either assume that the MSN is a graph or have studied the

|  |  |
| --- | --- |
| VOLUME x, 20xx | 13 |

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3009691, IEEE Access



Author et al.: Preparation of Papers for IEEE TRANSACTIONS and JOURNALS

mobility of individual users, therefore, most datasets are network graphs with attached user profiles or are trace sets.

**IX. FUTURE WORK**

In recent years, privacy in MSNs has gained a lot of atten-tion and has become an emerging field for research. MSNs are now part of a much larger paradigm of the Internet of Things (IoT)), therefore, they have formed more widespread networks. There is a lot of research being conducted currently to handle the enormous volume of data being generated by mobile devices in IoT, but the key point to note is that, with these advancements, more and more data is becoming available to attackers, allowing them to design more sophis-ticated attacks. Thus, there is a need for researchers to design privacy-preserving solutions that consider multiple sources of data over a single source, that can handle multiple data formats and can resist several attacks effectively. Following are the challenges and possible future research directions in privacy in MSNs :

Most of the data available publicly only anonymizes the user information and not their attribute information. These attributes have metadata that can be used to trace back to the user and this poses serious threats to the users. Therefore, it is crucial to resolve this.

Most privacy-preserving solutions consider a single server like LBS in their schemes. Though these solu-tions are effective now, they may not be compatible with newer standards like 5G, where we have hierarchical servers. Therefore, to design practical schemes that can withstand time, researchers can consider including these hierarchies, and the new network and cell organization.

In many real-world applications, the location privacy options are extreme and do not consider the user’s perceived utility of a check-in. Therefore, solutions can be designed based on the check-in’s context to the user. This makes the solutions more effective as a single location can have different privacy levels applied to it based on context, ensures better utility and meets user’s needs.

The privacy options are sometimes too complex for users to understand. Therefore, to relieve the user’s from making a decision, intelligent privacy mechanisms can be implemented to learn the user’s privacy preferences over time, so that in future instances, the privacy level is automatically applied.

Recently Reinforcement learning has seen a lot of ap-plications in privacy schemes in MSNs and VANETS, where RL has been used to select optimal privacy policies. Instead, RL can be used to study adversary behavior, which will give a better understanding of how an adversary works and this will be helpful in designing and fine-tuning better privacy strategies.

To deal with larger datasets and mine useful information for designing privacy schemes, advanced RL algorithms like policy gradient can be used.

As several sources of public data are available, solutions to try to address the issue of controlled linkability. Also, as the number of devices in MSNs increases trying to designing a centralized solution will increase the computation complexity. As MSNs have a decentral-ized/distributed aspect to them, lightweight anonymiza-tion solutions can be designed to anonymize data at the user end to greatly reduce the server load.

**ACKNOWLEDGEMENT**

This work was supported by NSF under Grant Nos. 1829674 and 1741277.

**REFERENCES**

1. "About GeoCities: Geocities Archive Geocities Mirror". [Online]. Avail-

able: http://www:geocitiesarchive:org

1. S. Phillips, "A brief history of Facebook", Jul., 2007. [Online]. Available: https://www:theguardian:com/technology/2007/jul/25/media:newmedia
2. W. Mitchell, "MySpace Rise and Fall and Rise-Again?". Jan., 2013. [Online]. Available: https://startupbros:com/myspace-the-rise-fall-and-rise-again-infographic/. [Accessed: Jun.-2018]
3. "About LinkedIn".[Online]. Available: https://about:linkedin:com
4. A.St. John, "Equifax Data Breach: What Consumers Need to Know",

25-Sept.-2017. [Online]. Available: shorturl:at/IJPSZ. [Accessed: 2-Oct.-2017]

1. Y. Chon, E. Talipov, H. Shin and H. Cha, "CRAWDAD dataset yonsei/lifemap", 3-Jan.-2012. [Online]. Available: https://crawdad:org/ yonsei/lifemap/20120103 , https://doi:org/10:15783/C7RW26.
2. D. Ingram, "Facebook says data leak hits 87 million users, widening pri-vacy scandal", Apr., 2018 .[Online]. Available: https://www:reuters:com/ article/us-facebook-privacy-idUSKCN1HB2CM
3. "Laplace distribution", Encyclopedia of Mathematics, Aug., 2014. [Online]. Available: http://encyclopediaofmath:org/index:php?title= Laplace\_distribution&oldid=33035
4. E. Rafaloff, "Shamir Secret Sharing Scheme", Oct., 2018. [Online]. Available: https://ericrafaloff:com/shamirs-secret-sharing-scheme/
5. I. Illascu, "Targeted Attacks : User Profiling", Sept., 2014. [Online].

Available: https://news:softpedia:com/news/User-Profiling-Services-

Could-Be-Compromised-for-More-Targeted-Attacks-460284:shtml

1. M. Gill, "Geo-Spoofing Explained", Jan., 2020. [Online]. Available: https://www:comparitech:com/blog/vpn-privacy/geospoofing/
2. E. Limer, "Eavesdrop through Phone Microphones to Target Ads", Jan., 2018. [Online]. Available: https://www:popularmechanics:com/ technology/security/a14533262/alphonso-audio-ad-targeting/
3. T. Kohut, "Our phones are listening to us", Feb., 2018. [Online].

Available: https://globalnews:ca/news/4039276/smart-devices-facebook-listening/

1. A. Langone, "How Facebook or Any Other App Could Use Your Phone’s Microphone to Gather Data", Mar., 2018. [Online]. Available: http://money:com/money/5219041/how-to-turn-off-phone-microphone-facebook-spying/
2. M. Kumar, "WireX DDoS Botnet", Aug., 2017. [Online]. Available: https://thehackernews:com/2017/08/android-ddos-botnet:html
3. A. Davies, "Akamai’s verdict on Mirai", Nov., 2016. [Online]. Available: https://disruptive:asia/akamai-assesses-mirai-bad-worse/
4. E. Tannam, "Phishing ad claiming to be Twitter verification ser-vice", May., 2018. [Online]. Available: https://www:siliconrepublic:com/ enterprise/twitter-ad-verification.
5. "The Rising Threat of Social Media Phishing Attacks", Oct., 2018. [On-line]. Available: https://corrata:com/the-rising-threat-of-social-media-phishing-attacks/
6. R. Dumont, "Adidas “prize” used as bait in attempt to lure people into biting", Jun. 2018. [Online]. Available: https://www:welivesecurity:com/ 2018/06/14/phishing-anniversary-free-50-month-subscription/
7. S. Gallagher, "Syrian rebels lured into malware honeypot sites", Feb., 2015. [Online]. Available: https://arstechnica:com/information-technology/2015/02/syrian-rebels-lured-into-malware-honeypot-sites-through-sexy-online-chats/

14 VOLUME x, 20xx

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3009691, IEEE Access



Author et al.: Preparation of Papers for IEEE TRANSACTIONS and JOURNALS

1. N. Kovacs, "Social Media Scams Based on Current Events", Aug., 2014. [Online]. Available: https://community:norton:com/blogs/norton-protection-blog/social-media-scams-based-current-events
2. L. Sweeney, "Simple Demographics Often Identify People Uniquely" in Health, vol. 671, pp. 1–34, 2000.
3. Z. Cai and X. Zheng, "A Private and Efficient Mechanism for Data Uploading in Smart Cyber-Physical Systems," in IEEE Transactions on Network Science and Engineering, vol. 7, no. 2, pp. 766-775, 2nd Quart., 2020, doi: 10:1109/TNSE:2018:2830307:
4. Y. Liu, W. Wei, A. Sun and C. Miao, "Exploiting geographical neigh-borhood characteristics for location recommendation", in Proceedings of the 23rd ACM International Conference on Information and Knowledge Management, pp. 739-748, Nov. 2014.
5. Y. Zheng, L. Zhang, X. Xie and W. Y. Ma, "Mining interesting locations and travel sequences from GPS trajectories", in Proceedings of the 18th international conference on World wide web, pp. 791-800, Apr. 2009.
6. Y. Zheng, Q. Li, Y. Chen, X. Xie, X. and W. Y. Ma, "Understanding mobility based on GPS data", in Proceedings of the 10th international conference on Ubiquitous computing, pp. 312-321, Sept. 2008.
7. Y. Chon, E. Talipov, H. Shin and H. Cha, "Mobility prediction-based smartphone energy optimization for everyday location monitoring", in Proceedings of the 9th ACM conference on embedded networked sensor systems pp. 82-95, Nov. 2011.
8. J. Chon and H. Cha, "Lifemap: A smartphone-based context provider for location-based services" in IEEE Pervasive Computing, vol. 10, no. 2, pp.58-67, 2011.
9. Y. Chon, H. Shin, E. Talipov and H. Cha, "Evaluating mobility models for temporal prediction with high-granularity mobility data", in 2012 IEEE International Conference on Pervasive Computing and Communi-cations,pp. 206-212), Mar. 2012.
10. A. R. Benson, D. F. Gleich and L. H. Lim, "The spacey random walk: A stochastic process for higher-order dat", SIAM Review, vol. 59, no. 2, pp. 321-345, 2017.
11. H. Li, H. Zhu, S. Du, X. Liang and X. Shen, "Privacy Leakage of Location Sharing in Mobile Social Networks: Attacks and Defense", IEEE Transactions on Dependable and Secure Computing, vol. 15, no. 4, pp. 646-660, Jul. 2018.
12. M. Li, H. Zhu, Z. Gao, S. Chen, L. Yu, S. Hu and K. Ren, "All your location are belong to us: Breaking mobile social networks for automated user location tracking", in Proceedings of the 15th ACM international symposium on Mobile ad hoc networking and computing, pp. 43-52, Aug. 2014.
13. N. Talukder and S. I. Ahamed, "Preventing multi-query attack in location-based services", in Proceedings of the third ACM conference on Wireless network security, pp. 25-36, Mar. 2010.
14. B. Lee, J. Oh, H. Yu and J. Kim, "Protecting location privacy using loca-tion semantics", in Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 1289-1297, Aug. 2011.
15. C. Y. Chow and M. F. Mokbel, "Enabling private continuous queries for revealed user locations", in International Symposium on Spatial and Temporal Databases, pp. 258-275, Jul. 2007.
16. J. Krumm, "Inference attacks on location tracks", International Confer-ence on Pervasive Computing, Springer, pp. 127-143, 2007.
17. Y. C. Hu, A. Perrig and D. B. Johnson, "Wormhole attacks in wireless networks", IEEE journal on selected areas in communications, vol. 24, no. 2, pp. 370 -380, Feb. 2006.
18. Y. A. De Montjoye, C. A. Hidalgo, M. Verleysen and V. D. Blondel, "Unique in the crowd: The privacy bounds of human mobility", Scientific reports, vol. 3, pp. 1376, 2013.
19. M. Maouche, S. B. Mokhtar and S. Bouchenak, "Ap-attack: a novel user re-identification attack on mobility datasets", in Proceedings of the 14th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, pp. 48 - 57, Nov. 2017.
20. K. Zhang, X. Liang, R. Lu and X. Shen, "Sybil Attacks and Their Defenses in the Internet of Things", IEEE Internet of Things Journal, vol. 1, no. 5, pp. 372 - 383, Oct. 2014.
21. R. John, J. P. Cherian and J. J. Kizhakkethottam, "A survey of tech-niques to prevent sybil attacks", 2015 International Conference on Soft-Computing and Networks Security (ICSNS), pp. 1-6, Feb. 2015.
22. G. Wang, T. Konolige, C. Wilson, X. Wang, H. Zheng and B. Y. Zhao, "You are how you click: Clickstream analysis for sybil detection", in 22nd USENIX Security Symposium (USENIX Security 13), pp. 241-256, 2013.

VOLUME x, 20xx

1. G. U. Vasanthakumar, K. Sunithamma, P. D. Shenoy and K. R. Venu-gopal, "An overview on user profiling in online social networks", Int. J. Appl. Inf. Syst, vol. 11, no. 8, pp. 25- 42, 2017.
2. H. Li, H. Zhu, S. Du, X. Liang and X. Shen, "Privacy Leakage of Location Sharing in Mobile Social Networks: Attacks and Defense", in IEEE Transactions on Dependable and Secure Computing, vol. 15, no. 4, pp. 646 - 660, Jul. 2018.
3. X. Zheng, G. Luo and Z. Cai, "A Fair Mechanism for Private Data Pub-lication in Online Social Networks", in IEEE Transactions on Network Science and Engineering, pp.1-1, Feb. 2018.
4. S. Zhang, G. Wang, Q. Liu, Q and J. H. Abawajy, "A trajectory privacy-preserving scheme based on query exchange in mobile social networks", Soft Computing, Springer, vol. 22, no. 18, pp. 6121-6133, 2018.
5. K. Huguenin, I. Bilogrevic, J. S. Machado, S. Mihaila, R. Shokri, I. Da-costa and J. Hubaux, "A Predictive Model for User Motivation and Utility Implications of Privacy-Protection Mechanisms in Location Check-Ins", in IEEE Transactions on Mobile Computing, vol. 17, no. 4, pp. 760-774, Apr. 2018.
6. X. Xiao, C. Chen, A. K. Sangaiah, G. Hu, R. Ye and Y. Jiang, "Cen-LocShare: a centralized privacy-preserving location-sharing system for mobile online social networks", Future Generation Computer Systems, vol. 86, pp. 863-872, Feb. 2017.
7. T. N. Phan, T. K. Dang, T. A. Truong and T. H. Lam, "A Context-Aware Privacy-Preserving Solution for Location-Based Services", in 2018 International Conference on Advanced Computing and Applica-tions (ACOMP), pp. 132-139, Nov. 2018.
8. T. Peng, Q. Liu, D. Meng and G. Wang, "Collaborative trajectory privacy preserving scheme in location-based services", Information Sciences, vol. 387, pp. 165-179, Aug. 2016.
9. X. Zheng, Z. Cai, J. Yu, C. Wang and Y. Li, "Follow But No Track: Privacy Preserved Profile Publishing in Cyber-Physical Social Systems", in IEEE Internet of Things Journal, vol. 4, no. 6, pp. 1868-1878, Dec. 2017.
10. X. Zheng, Z. Cai and Y. Li, "Data Linkage in Smart Internet of Things Systems: A Consideration from a Privacy Perspective", in IEEE Commu-nications Magazine, vol. 56, no. 9, pp. 55-61, Sep. 2018.
11. Y. Liang, Z. Cai, J. Yu, Q. Han and Y. Li, "Deep Learning Based Inference of Private Information Using Embedded Sensors in Smart Devices", in IEEE Network, vol. 32, no. 4, pp. 8-14, Jul. 2018.
12. Z. Cai, Z. He, X. Guan and Y. Li, "Collective Data-Sanitization for Preventing Sensitive Information Inference Attacks in Social Networks", in IEEE Transactions on Dependable and Secure Computing, vol. 15, no. 4, pp. 577-590, Jul. 2018.
13. R. Dewri, "Local Differential Perturbations: Location Privacy under Approximate Knowledge Attackers", in IEEE Transactions on Mobile Computing, vol. 12, no. 12, pp. 2360-2372, Dec. 2013.
14. C. Yin, J. Xi, R. Sun and J. Wang, "Location Privacy Protection Based on Differential Privacy Strategy for Big Data in Industrial Internet of Things", in IEEE Transactions on Industrial Informatics, vol. 14, no. 8, pp. 3628-3636, Aug. 2018.
15. W. Wang, M. Min, L. Xiao, Y. Chen and H. Dai, "Protecting Semantic Trajectory Privacy for VANET with Reinforcement Learning", 2019 IEEE International Conference on Communications (ICC), pp. 1-5, 2019.
16. M. Li, N. Cao, S. Yu and W. Lou, "Findu: Privacy-preserving personal profile matching in mobile social networks", in 2011 Proceedings IEEE INFOCOM, pp. 2435-2443, Apr. 2011.
17. R. Li, H. Li, X. Cheng, X. Zhou, K. Li, S. Wang and R. Bie, "Perturbation-based private profile matching in social networks", IEEE Access, vol. 5, pp. 19720-19732, Sep. 2017.
18. F. Li, Y. He, B. Niu, H. Li and H. Wang, "Match-MORE: An efficient private matching scheme using friends-of-friends’ recommendation", 2016 International Conference on Computing, Networking and Commu-nications (ICNC), pp. 1-6, Feb. 2016.
19. A. Campan and T. M. Truta, "A clustering approach for data and structural anonymity in social networks", in Proceeding of the second ACM SIGKDD International Workshop on Privacy, Security, and Trust in KDD, Jun. 2008.
20. M. Siddula, Z. Cai and D. Miao, "Privacy Preserving Online Social Networks using Enhanced Equicardinal Clustering", in 2018 IEEE 37th International Performance Computing and Communications Conference (IPCCC), pp. 1-8, Nov. 2018.
21. A. Pingley, W. Yu, N. Zhang, X. Fu and W. Zhao, "A context-aware scheme for privacy-preserving location-based services", Computer Net-works, vol. 56. no. 11, pp. 2551-2568, Jul. 2012.

15

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3009691, IEEE Access



Author et al.: Preparation of Papers for IEEE TRANSACTIONS and JOURNALS

1. S. Zhang, G. Wang, M. Z. A. Bhuiyan and Q. Liu, "A Dual Privacy Preserving Scheme in Continuous Location-Based Services", in IEEE Internet of Things Journal, vol. 5, no. 5, pp. 4191-4200, Oct. 2018.
2. M. Hazazi, Y. Tian and M. Al-Rodhaan, "Privacy-Preserving Authenti-cation Scheme for Wireless Networks", in 2018 21st Saudi Computer Society National Computer Conference (NCC), pp. 1-6, Apr. 2018.
3. L. Zhu, C. Zhang, C. Xu, X. Du, R. Xu, K. Sharif and M. Guizani, "PRIF: A Privacy-Preserving Interest-Based Forwarding Scheme for Social Internet of Vehicles", in IEEE Internet of Things Journal, vol.5, no.4, pp. 2457-2466, Aug. 2018.
4. P. Syverson, D. Goldschlag and M. Reed, "Onion routing for anonymous and private internet connections", Communications of the ACM, vol. 42, no. 2, Feb. 1999.
5. I. Aad, C. Castelluccia and J. P. Hubaux, "Packet coding for strong anonymity in ad hoc networks", in 2006 Securecomm and Workshops,
   1. 1-10, Aug. 2006.
6. O. Hasan, J. Miao, S. B. Mokhtar and L. Brunie, "A Privacy Preserving Prediction-based Routing Protocol for Mobile Delay Tolerant Networks", in 2013 IEEE 27th International Conference on Advanced Information Networking and Applications (AINA), pp. 546-553, Mar. 2013.
7. C. Y. Chow, "Cloaking Algorithms for Location Privacy", Encyclopedia of GIS, Springer US, pp. 93 -97, 2008, doi:10.1007/978-0-387-35973-1\_136.
8. P. Wightman, W. Coronell, D. Jabba, M. Jimeno and M. Labrador, "Eval-uation of location obfuscation techniques for privacy in location based information systems", in 2011 IEEE Third Latin-American Conference on Communications, pp. 1-6, Oct. 2011.
9. C. Y. Chow, M. F. Mokbel and X. Liu, "Spatial cloaking for anonymous location-based services in mobile peer-to-peer environments", GeoInfor-matica, vol. 15, no. 2, pp. 351-380, Apr. 2011, doi:10.1007/s10707-009-0099-y.
10. M. Gruteser and D. Grunwald, "Anonymous usage of location-based services through spatial and temporal cloaking", in Proceedings of the 1st international conference on Mobile systems, applications and services,
    1. 31-42, Jan. 2003.
11. N. Mattani, J. S. Kumar, A. Prabakaran and N. Maheswari, "Privacy Preservation in Social Network Analysis using Edge Weight Perturba-tion", Indian Journal of Science and Technology, vol. 9, no. 37, pp. 1-10, Oct. 2018.
12. A. Campan and T. M. Truta, "Data and structural k-anonymity in social networks. In International Workshop on Privacy", Security and Trust in KDD, Springer, vol. 5456, pp. 33-54, Aug. 2009.
13. S. Jahid, P. Mittal and N. Borisov, "EASiER: Encryption-based access control in social networks with efficient revocation", in Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security, pp. 411-415, Mar. 2011.
14. K. El Emam, E. Jonker, L. Arbuckle and B. Malin, "A systematic review of re-identification attacks on health data", PloS one, vol. 6, no. 12, pp. e28071, Dec. 2011.
15. J. Leskovec and A. Krevl, "SNAP Datasets: Stanford Large Network

Dataset Collection, (2014). [Online]. Available: http://snap:stanford:edu/ data.

1. Facebook Dataset. [Online]. Available: https://snap:stanford:edu/data/

ego-Facebook:html.

1. Twitter Dataset. [Online]. Available: https://snap:stanford:edu/data/

ego-Twitter:html

1. B. Rozemberczki, R. Davis, R. Sarkar and C. Sutton, "Gemsec: Graph

embedding with self clustering" (2018). [Online]. Available: https://

arxiv:org/abs/1802:03997.

1. Deezer Dataset, [Online]. Available: https://snap:stanford:edu/data/

gemsec-Deezer:html

1. E. Cho, S.A. Myers, and J. Leskovec, "Friendship and mobility: user movement in location-based social networks", in Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, Aug., 2018, pp. 1082-1090.
2. E. Cho and S. A. Myers and J. Leskovec, "Brightkite Dataset" . [Online]. Available: https://snap:stanford:edu/data/loc-Brightkite:html
3. Gowalla Dataset. [Online]. Available: http://www:yongliu:org/datasets/

index:html

1. R. Zafarani and H. Liu, Social Computing Data Repository at ASU (2009). [Online]. Available: http://socialcomputing:asu:edu
2. Geocities Dataset. [Online]. Available: https://www:microsoft:com/en-

us/download/details:aspx?id=52367

1. Manhattan Taxi trajectory Dataset. [Online]. Available: https://

www:cs:cornell:edu/~arb/data/Manhattan-taxi-trajectories/index:html

1. C. Whong, "FOILing NYC’s Taxi Trip Data". [Online]. Available: https:

//chriswhong:com/open-data/foil\_nyc\_taxi/

1. M. Piorkowski, N.Sarafijanovic-Djukic, and M.Grossglauser, "CRAW-DAD data set epfl/mobility "(2009). [Online]. Available: https:// crawdad:org/epfl/mobility/20090224/cab
2. T-Drive Dataset. [Online]. Available: https://privamov:github:io/accio/

docs/datasets:html#t-drive

1. Mackaroo - online data generator. [Online]. Available: https://

mockaroo:com

1. Generatedata - online data generator. [Online]. Available: http://

www:generatedata:com

1. DTMGenerator - online data generator. [Online]. Available: https://

www:rankred:com/test-data-generation-tools/

AKSHITA MARADAPU VERA VENKATA SAI (S’18) received the B.S. degree in computer sci-ence and engineering from GITAM University, India, in 2016, and the M.S. degree in computer science from Georgia State University, USA, in 2017, where she is currently pursuing the Ph.D. degree with the Department of Computer Science. Her research interests include mobile social net-works, security and privacy, deep learning, and big data mining.



YINGSHU LI (SM’10) received the B.S. de-gree from the Department of Computer Science and Engineering, Beijing Institute of Technology, China, and the M.S. and Ph.D. degrees from the Department of Computer Science and Engi- neer-ing, University of Minnesota–Twin Cities. She is currently a Professor with the Department of Computer Science, Georgia State University. Her research interests include wireless networking, sensor networks, sensory data management, social



networks, and optimization. She was a recipient of the NSF CAREER Award.



16 VOLUME x, 20xx

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.