Differential Privacy and Prediction Uncertainty of Gossip Protocols in General Networks

Yufan Huang, Richeng Jin and Huaiyu Dai
Department of Electrical and Computer Engineering
North Carolina State University, Raleigh, USA
{yhuang20,rjin2,hdai}@ncsu.edu

***Abstract*—Recent advances in social media and information technology have enabled much faster dissemination of information, while at the same time raise concerns about privacy leakage after various privacy breaches. Therefore, the privacy guarantees of information dissemination protocols have attracted increasing research interests, among which the gossip protocols assume vital importance in various information exchange applications. Very recently, the rigorous framework of differential privacy has been introduced to measure the privacy guarantees of gossip protocols in the simplified complete network scenario. In this work, we extend the study to general networks. First, lower bounds of the differential privacy guarantees are derived for the gossip protocols in general networks in both synchronous and asynchronous settings. The prediction uncertainty of the source node given a uniform prior is also determined. It is found that source anonymity is closely related to some key network structure parameters in the general network setting. Then, we investigate information spreading in wireless networks with unreliable communications, and quantify the tradeoff between differential privacy guarantees and information spreading efficiency. Finally, considering that the attacker may not be present in the beginning of the information dissemination process, the scenario of delayed monitoring is studied and the corresponding differential privacy guarantees are evaluated.**

I. Introduction

It is well-known that most people are six or fewer social connections away from each other. Recently, the explosive development in the Internet and social networks makes it easy for people to disseminate their information to the rest of the world. Gossip protocols, in which networked nodes randomly choose a neighbor to exchange information, have been widely adopted in various applications for information dissemination due to their simplicity and efficiency. For instance, they can be used to spread and aggregate information in dynamic networks like mobile networks, wireless sensor networks, and unstructured P2P networks [1]–[3]. Combined with stochastic gradient descent methods, gossip protocols are also adapted to implement distributed machine learning [4], [5]. In particular, the authors of [5] propose to transmit differentially private gradient information through gossip protocols. Nonetheless, they focus on the privacy of the shared gradient information rather than the anonymity of the source.

With the arising concerns of privacy exposure, the information sources often prefer to stay anonymous while disseminating some sensitive information. Gossip protocols are believed to provide a certain form of source anonymity since most nodes don't get informed directly from the source, and the origin of the information becomes increasing blurred as the spreading proceeds. In this regard, source identification and protection of gossip protocols have attracted significant research interests (see [6], [7] and the references therein). However, the existing approaches usually assume some specific network structures (e.g., tree graphs) and attacking techniques (e.g., maximum likelihood estimator) and don't easily generalize.

To study the privacy of gossip protocols in a formal and rigorous setting, the concept of differential privacy [8], which was originally introduced in data science, is adapted to measure the source anonymity of gossip protocols in [9]. However, their study is restricted to complete networks, which may not be a good model in practice. For example, practical networks often have a network diameter much larger than 1 (41 for the Facebook network [10]).

In this work, we extend the study of the fundamental limits on the privacy of gossip-based information spreading protocols to general networks. Our main contributions are summarized as follows.

1) Lower bounds of the differential privacy guarantees of general gossip protocols are derived for general networks in both synchronous and asynchronous settings. The prediction uncertainty of the source node given a uniform prior is also determined.
2) The differential privacy of standard gossip and private gossip protocols is further studied in a wireless setting, where communications are assumed to be unreliable. It is found that wireless interference can enhance the differential privacy while slowing down the spreading process. Through analysis and simulations, the tradeoff between the differential privacy guarantees and the information spreading efficiency is revealed.
3) Finally, the effect of the additional uncertainty induced by delayed monitoring on the differential privacy guarantees is shown.

II. System Model

A. Gossip Protocol

In this work, we investigate the privacy of information source in gossip-based information spreading. The goal is to measure

This work was supported in part by the US National Science Foundation under grants ECCS-1444009 and CNS-1824518, and in part by US Army Research Office under Grant W911NF-17-1-0087.

the capability of gossip protocols in keeping the information source anonymous. Specifically, given a connected network $G = (V, E)$ of arbitrary topology, where $V = \{0, 1, ..., n-1\}$ is the node set and E is the set of connecting edges, a node (source) initially possesses a piece of information and needs to deliver it to all the other nodes in the network. All the nodes are assumed to share the same communication protocol *gossip*. Each time an informed node i performs *gossip*, it will contact one of its neighboring nodes $j \in N_i$ with probability $1/d_i$, where d_i is the degree of node i. The whole information dissemination process terminates after all the nodes are informed. Same as [9], we focus on the gossip protocols based on the "push" action[1] in this work, and consider the following two specific gossip protocols.

1) Standard Gossip: All informed nodes remain active (i.e., continuously performing *gossip*) during the spreading process.
2) Private Gossip [9]: Once an active informed node (initially it is the source) performs gossip, it turns inactive and the newly informed node takes over the source role.

B. Time Model

Both synchronous and asynchronous time models are considered. In the former, all nodes share a global discrete time clock. Each time the clock ticks, all active informed nodes perform the *gossip* action simultaneously, and the informed node set is updated accordingly, counted as one round. In the asynchronous time model, each node has its own internal clock, which ticks according to a Poisson process, with the mean interval between two ticks equivalent to that of one round in the synchronous model. The gossip action and update of the informed node set is performed each time the clock of an active informed node ticks.

C. Threat Model

The goal of the attacker is to identify the source node based on its observations (i.e., attack on confidentiality and privacy). It is assumed that the attacker can monitor the ongoing communications in the whole network, through, e.g., deploying a sufficient number of sensors throughout the field. With a probability of $0 < \alpha \leq 1$, the sensors can correctly observe the identities of the active nodes at each gossip step. Specifically, as shown in Fig. 1, the observed event has the form of $\mathcal{S} = ((i, t)), i \in V, t \in \{0, 1, 2, \cdots\}$ in the synchronous setting, which indicates that the attacker knows node i performs the gossip action at time slot t. In the asynchronous setting, however, the attacker does not know the exact time of each observed event, but only the relative order of the nodes' activities. The observed event in this case is represented by $\mathcal{S} = ((i|t))$, where the condition t stands for the latent time information unknown to the attacker.

[1]In the corresponding "pull" action, uninformed nodes are active and try to solicit the information from informed nodes. The "push" action is dominant for information spreading in social and mobile networks. In addition, such a study is also conservative in the sense that it gives the attacker an advantage by only monitoring the "push" actions.

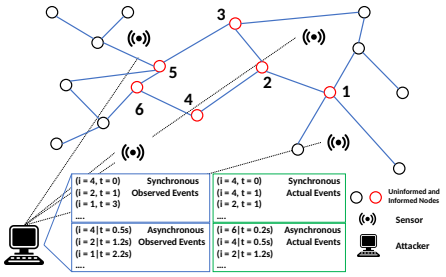

Fig. 1: Sensor Monitoring and Observations

D. Privacy Model

In this work, differential privacy is adopted to measure the information leakage of the gossip protocols. In particular, a randomized algorithm $\mathcal{R}$ with domain $\mathbb{N}^{|\mathcal{X}|}$ is (ϵ, δ)-differentially private if for all $\mathcal{S} \subseteq Range(\mathcal{R})$ and for any two databases x, y that differ on a single element [8]:

$$Pr[\mathcal{R}(x) \in \mathcal{S}] \leq e^{\epsilon} Pr[\mathcal{R}(y) \in \mathcal{S}] + \delta, \tag{1}$$

where parameter $\epsilon \geq 0$ is the privacy budget while $\delta \geq 0$ is the tolerance level for the violation of the ϵ bound. Specifically, given the privacy budget ϵ and the tolerance level δ, Eq. (1) implies that the randomized algorithm guarantees that the privacy loss is bounded by ϵ with a probability of at least $1 - \delta$. Consider a source indicator database of the format $\mathcal{D}^{(i)} = [0, ..., d_i = 1, ..., 0]$ with exactly one non-zero value $d_i = 1$ if node i is the source. Given $D \triangleq \{\mathcal{D}^{(i)}\}_{i=0}^{n-1}$ and the graph G as the input, a gossip protocol can be treated as a randomized algorithm with the output set $\mathbb{S}$ (i.e., the range) consisting of all possible observation sequences by the attacker during the execution of the protocol.

Definition 1: Given a general network G, a gossip protocol is (ϵ, δ)-differentially private in G if for all observations $\mathcal{S} \subseteq \mathbb{S}$ and for any two source indicator vectors $\mathcal{D}^{(i)}, \mathcal{D}^{(j)}, i, j \in V$:

$$p_G^{(i)}(\mathcal{S}) \leq e^{\epsilon} p_G^{(j)}(\mathcal{S}) + \delta, \tag{2}$$

where $p_G^{(i)}(\mathcal{S}) = Pr[\mathcal{S}|G, \mathcal{D}^{(i)}]$ is the conditional probability of an observation event $\mathcal{S}$ given the network graph G and the source indicator vector $D^{(i)}$.

In this work, considering the fact that, due to the topological and observation model constraints, there may exist some (rare) events $\mathcal{S}$ such that $p_G^{(j)}(\mathcal{S}) = 0$ (e.g., if $\mathcal{S}_{i,0}$ is the observed event that node i performs gossip at time 0 in the synchronous setting, then $p_G^{(j)}(\mathcal{S}_{i,0}) = 0, \forall j \neq i$), additional tolerance level is needed to ensure the privacy guarantees (i.e., the pure version of the differential privacy with $\delta = 0$ is infeasible). Thus, this study will mainly focus on the tolerance level δ of the privacy guarantees provided by the gossip protocols for any given privacy budget ϵ. Clearly, $\delta \leq 1$ for any $\epsilon \geq 0$.

In addition to differential privacy, it is also desirable to study privacy guarantees of information dissemination protocols from a more pertinent perspective, i.e., source identification through prediction or detection. Reusing the above example, there always exist some events $\mathcal{S}$ such that $p_G^{(i)}(\mathcal{S}) > 0$ for some i but $p_G^{(j)}(\mathcal{S}) = 0, \forall j \neq i \in V$, which satisfy an arbitrary privacy budget ϵ with a tolerance level of

δ (if $p_G^{(i)}(\mathcal{S}) \leq \delta$). However, the identity of the source (i.e., node i) can still be easily inferred. Therefore, it is further required that some *prediction uncertainty* be guaranteed for a given differentially private protocol, which is defined as [9]:

Definition 2: Given a general network G, the prediction uncertainty of a gossip protocol is defined for a uniform prior $p_G(I_0)$ on source nodes and any $i \in \{0, 1, ..., n-1\}$ as:

$$c = \min_{i, \mathcal{S} \subseteq \mathbb{S}} \left(\frac{p_G(I_0 \neq \{i\}|\mathcal{S})}{p_G(I_0 = \{i\}|\mathcal{S})} \right), \forall p_G^{(i)}(S) > 0, \tag{3}$$

where I_0 stands for the initial informed node set and its element represents the source node.

Remark 1: The connection of prediction uncertainty and differential privacy is illustrated below. If the attacker obtains an observation $\mathcal{S}$, differential privacy measures the probabilities of it observing $\mathcal{S}$ given different sources while prediction uncertainty considers the posterior probabilities of different sources given $\mathcal{S}$. Especially, because of the uniform prior $p_G(I_0)$, $\frac{p_G(I_0 \neq \{i\}|\mathcal{S})}{p_G(I_0 = \{i\}|\mathcal{S})} = \frac{\sum_{j \neq i} p_G^{(j)}(\mathcal{S})}{p_G^{(i)}(\mathcal{S})}$ holds by the Bayes' formula. Prediction uncertainty is an appealing metric in this study as it measures the privacy guarantees from the source prediction perspective with a much smaller cardinality than the classic privacy budget (which requires the study of all pairs of $p_G^{(i)}(\mathcal{S})$ and $p_G^{(j)}(\mathcal{S})$). Moreover, given a prediction uncertainty c, it can be shown that $p_G(I_0 = \{i\}|\mathcal{S}) \leq \frac{1}{c+1}, \forall i, \mathcal{S}$; therefore a larger c indicates better source anonymity.

III. Main Results

To facilitate our following analysis, we need the following lemma and definition of *decay centrality*.

Lemma 1: Given any gossip protocol in a graph G, let $\mathcal{S} \subseteq \mathbb{S}$ and there are two constants $w_G^{(i)}(\mathcal{S}), w_G^{(j)}(\mathcal{S})$ such that $p_G^{(i)}(\mathcal{S}) \geq w_G^{(i)}(\mathcal{S})$ and $p_G^{(j)}(\mathcal{S}) \leq w_G^{(j)}(\mathcal{S})$. If the gossip protocol satisfies (ϵ, δ)-differential privacy, then $\delta \geq \max_{\mathcal{S},i,j}(w_G^{(i)}(\mathcal{S}) - e^{\epsilon} w_G^{(j)}(\mathcal{S}))$.

Lemma 1 readily follows from the definition of differential privacy; its proof is omitted in the interest of space.

Definition 3: [11] Given a network G and a decay parameter β, $0 < \beta < 1$, the **decay centrality** of node i is defined as

$$C_\beta(i) = \sum_{j \neq i} \beta^{d(i,j)}, \tag{4}$$

where $d(i, j)$ is the length of the shortest path between node i and j.

Remark 2: Decay centrality measures the ease of a node reaching out to other nodes in the network. A large decay centrality indicates the central positioning of a node and its easiness to reach other nodes. The difficulty increases as β decreases.

A. Privacy of Gossip Protocols in General Networks

Our main result concerning the privacy guarantees of general gossip protocols in a general network is given below.

Theorem 1: Given a connected network G with n nodes and diameter $D_G = \max_{i,j \in V, i \neq j} d(i, j)$, and considering the observation model described in Section II-C with parameter α, if a gossip protocol satisfies (ϵ, δ)-differential privacy for any $\epsilon \geq 0$ and c-prediction uncertainty, then we have $\delta \geq \alpha$ and $c = 0$ in the synchronous setting. In the asynchronous setting,

$$\delta \geq max[\alpha - e^{\epsilon}(1-\alpha)^{D_G}, \alpha - e^{\epsilon}\frac{1-\alpha}{n-1}] \tag{5}$$

and

$$c \leq \min_{i \in V} \frac{C_{1-\alpha}(i)}{\alpha}, \tag{6}$$

where $C_{1-\alpha}(i)$ is the decay centrality of node i with decay parameter $1 - \alpha$.

Sketch of proof: First, for the synchronous setting, let $\mathcal{S}_{i,0}$ be the event that node i's activity is observed by the attacker's sensors at time 0. Then, the probability that such an event happens given the source node is i is $p_G^{(i)}(\mathcal{S}_{i,0}) = \alpha$. If the source node is any other node $j \neq i$, $p_G^{(j)}(\mathcal{S}_{i,0}) = 0$ since node i cannot initialize a communication if it is not a source node at time 0. Therefore, $\delta \geq \alpha$ and $c = 0$.

In the asynchronous setting, let $\mathcal{S}_{i,0}$ be the event that node i's activity is observed by the attacker's sensors as its first observed event. It can be seen that, if the source node is i, then $p_G^{(i)}(\mathcal{S}_{i,0}) = p_G^{(i)}(\mathcal{S}_{i,0}|T_{i,0})p_G^{(i)}(T_{i,0}) + p_G^{(i)}(\mathcal{S}_{i,0}|\overline{T}_{i,0})p_G^{(i)}(\overline{T}_{i,0}) \geq \alpha$, where $T_{i,0}$ stands for the event that the source node is detected during its first communication. If the source node is j, we can consider the following event, denoted as $O_{d(i,j)}$: there is no communication detected by the sensors in the network after $d(i, j)$ gossip actions have been executed from the beginning. Then we have

$$\begin{aligned} p_G^{(j)}(\mathcal{S}_{i,0}) &= p_G^{(j)}(\mathcal{S}_{i,0} \bigcap \overline{O}_{d(i,j)}) + p_G^{(j)}(\mathcal{S}_{i,0} \bigcap O_{d(i,j)}) \\ &= p_G^{(j)}(\mathcal{S}_{i,0} \bigcap O_{d(i,j)}) \\ &\leq p_G^{(j)}(O_{d(i,j)}) = (1-\alpha)^{d(i,j)}, \end{aligned} \tag{7}$$

where the second equality is due to the fact that $\mathcal{S}_{i,0} \bigcap \overline{O}_{d(i,j)} = \emptyset$, as it takes at least $d(i, j)$ communications for the information to be delivered to node i from node j. Since $p_G^{(i)}(\mathcal{S}_{i,0}) \geq \alpha$, by applying Lemma 1, we have

$$\delta \geq \max_{i,j}(\alpha - e^{\epsilon}(1-\alpha)^{d(i,j)}) = \alpha - e^{\epsilon}(1-\alpha)^{D_G}. \tag{8}$$

On the other hand, since $\sum_{j \in V} p_G^{(i)}(\mathcal{S}_{j,0}) = 1$, there exists a node $l \in V$ such that

$$\begin{aligned} p_G^{(i)}(\mathcal{S}_{l,0}) &\leq \frac{1}{n-1} \sum_{j \in V, j \neq i} p_G^{(i)}(\mathcal{S}_{j,0}) \\ &= \frac{1 - p_G^{(i)}(\mathcal{S}_{i,0})}{n-1} \leq \frac{1-\alpha}{n-1}. \end{aligned} \tag{9}$$

This implies $\delta \geq \alpha - e^{\epsilon}\frac{1-\alpha}{n-1}$. By Eq. (8), we have

$$\delta \geq max[\alpha - e^{\epsilon}(1-\alpha)^{D_G}, \alpha - e^{\epsilon}\frac{1-\alpha}{n-1}]. \tag{10}$$

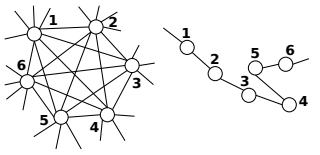

Fig. 2: Node 1 and node 6 are more distinguishable in the right network.

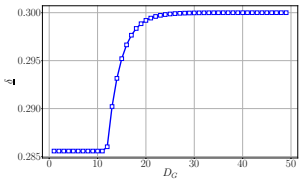

Fig. 3: Tolerance Level v.s. Network Diameter ($\alpha = 0.3, n = 50, \epsilon = 0.01$).

Meanwhile, as we have $p_G^{(j)}(\mathcal{S}_{i,0}) \leq (1-\alpha)^{d(i,j)}$, the detection uncertainty can be calculated as

$$c = \min_{i,\mathcal{S}}\left(\frac{\sum_{j\neq i} p_G^{(j)}(\mathcal{S})}{p_G^{(i)}(\mathcal{S})}\right) \leq \min_i \frac{\sum_{j\neq i}(1-\alpha)^{d(i,j)}}{\alpha} = \min_{i\in V} \frac{C_{1-\alpha}(i)}{\alpha}. \quad (11)$$

■

Remark 3: Some interpretations of the results of Theorem 1 are in order. It can be observed that the asynchronous setting provides better privacy guarantees than the synchronous setting, since the attacker has less information (i.e., the timing of the events) in this case. Note that differential privacy considers the worst case scenario. In the synchronous setting, when the attacker detects the activity of a node at time 0, it can infer that the corresponding node is the source immediately. Therefore, the prediction uncertainty is 0 due to this worst-case event, and the privacy guarantees are determined by the attacker's sensing capability α in the synchronous setting. In the asynchronous setting, however, the attacker could not directly infer the source solely based on the first-observed event due to the lack of associated timing. A counter example can be found in Fig. 1.

As a result, the structure of the network plays an important role in the asynchronous setting. In the context of information spreading, if two nodes are further apart, it takes more time for the information to be spread from one to the other; this duration gives the attacker more opportunities to differentiate the detected events, which leads to potentially higher privacy loss of the source node's identity. For instance, in the left network of Fig. 2, considering the event $\mathcal{S}_{1,0}$, i.e., node 1's activity being the first observed event by the attacker in the asynchronous setting, the probability of this event given that the source is 6 is $p_G^{(6)}(\mathcal{S}_{1,0}) \leq (1-\alpha)$ according to (7) and the probability of this event given that the source is 1 is $p_G^{(1)}(\mathcal{S}_{1,0}) \geq \alpha$. But in the right network, the corresponding probabilities are $p_G^{(6)}(\mathcal{S}_{1,0}) \leq (1-\alpha)^5$ and $p_G^{(1)}(\mathcal{S}_{1,0}) \geq \alpha$, which makes $\mathcal{S}_{1,0}$ a more distinguishable event in the right network. Therefore, the network diameter D_G, as the distance measure of the whole network, captures the potential privacy loss and becomes a key factor of the differential privacy lower bound in (5); an example of the relationship between the differential privacy tolerance level and the network diameter is shown in Fig. 3. The same logic is reflected on the prediction uncertainty given in (6). The smaller the decay centrality a network has (i.e., the nodes are more distant from each other), the more likely the attacker can identify the source node through its observations. Therefore, the inherent network structure imposes certain limit on privacy preserving concerning the source node identity, which applies to all information spreading protocols and calls for other privacy protection mechanisms, to be further explored in future work.

In addition, it can be seen that as the attacker's sensing capability α increases the privacy guarantees decrease (i.e., δ increases and c decreases). In particular, for an omnipresent attacker with $\alpha = 1$, we have $\delta = 1$ and $c = 0$ even in the asynchronous setting.

B. Privacy-Spreading Tradeoff of Gossip Protocols in Wireless Networks

Considering that in many real world applications, the information spreading between two nodes may be realized through wireless communications [2], [12], the privacy guarantees of gossip protocols in wireless networks are investigated in this subsection. It is assumed that the communications between the network nodes and between the attacker and its deployed sensors are prone to errors due to various interferences. To simplify the analysis, a failure probability is considered in this setting: Due to interferences, the communications will fail with a probability of f between two nodes during the $gossip$ step, and it is assumed that the attacker fails to receive a report from any of its deployed sensors about the detected events with the same probability f.[2] Note that the failure probability f, induced by detrimental effects in wireless channels, is different from the detection probability α that is due to the limitation in the eavesdropping capability (e.g., computation power) of the sensors. In this case, the privacy guarantees of gossip protocols are characterized in the following theorem.

Theorem 2: Considering the same setting as in Theorem 1, with the additional constraint that both the legitimate communication and the adversarial reporting fail with a probability f, the gossip-based protocols can guarantee (ϵ, δ)-differential privacy with $\delta \geq \alpha(1-f)$ and c-prediction uncertainty with $c = 0$ in the synchronous setting, and $\delta \geq max[\alpha(1-f) - e^{\epsilon}(1-\alpha(1-f))^{D_G}, \alpha(1-f) - e^{\epsilon}\frac{1-\alpha(1-f)}{n-1}]$ and $c \leq \min_{i\in V} \frac{C_{1-\alpha(1-f)}(i)}{\alpha(1-f)}$ in the asynchronous setting.

Sketch of Proof: This follows from the previous results and the details are omitted in the interest of space.

Adding artificial noise is a typical way to enhance privacy in practical applications [14]. In wireless networks, interference is a natural source for privacy enhancement as it hampers the attacker's observations of the network activities, which can be further strengthened through approaches such as friendly jamming [15]. However, the information spreading process is impeded as well in such scenarios. The information spreading time of the standard and the private gossip protocols in this case is given below.

[2]As a first work in this area, this simplified assumption is adopted to facilitate the characterization of the tradeoff between privacy and spreading speed. More realistic assumption concerning two different but correlated failure probabilities [13] warrants further study.

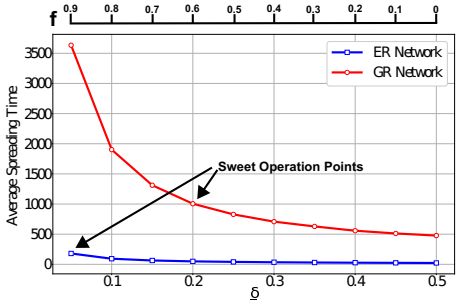

Fig. 4: DP v.s. spreading speed in the synchronous setting

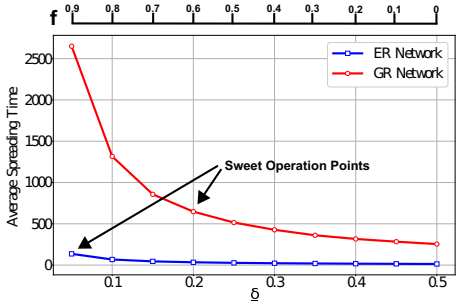

Fig. 5: DP v.s. spreading speed in the asynchronous setting

Theorem 3: In a wireless network G in which the communications fail with a probability of f, we have

1) In the synchronous setting, the private gossip takes $C_G/(1-f)$ rounds on average to inform all nodes in the network, where C_G is the cover time of a random walk in network G.
2) In the asynchronous setting, the private gossip takes $C_G/(1-f)$ time on average, while the standard gossip takes $T_{as}/(1-f)$ time on average to finish spreading, where T_{as} is the spreading time of standard gossip when the communication is perfect.

Sketch of Proof: Private gossip is a single random walk on the graph. The average time to inform all nodes is equal to C_G in both the synchronous and asynchronous settings. Given a failure probability of f, the interstate time is amplified by a factor of $1/(1-f)$. The same logic can be applied to standard gossip in the asynchronous setting.

Remark 4: For standard gossip in the synchronous setting, multiple random walks can exist during the spreading process, which renders the analysis of unreliable spreading challenging in general networks. But we conjecture that a similar result as in the asynchronous setting may hold.

The above results indicate a trade-off between privacy and spreading speed of gossip protocols, which is further explored through simulations below. In particular, following the existing literature in information spreading (e.g., [16], [17]), Erdős Rényi (ER) networks and Geometric Random (GR) Networks with a total number of $n = 100000$ nodes and average node degree of 10 are considered. Each point in the following figures is obtained through simulations with 5 network instances and 100 Monte Carlo runs for each instance. The average 90% spreading time is considered [12]. The privacy-spreading tradeoffs for ER and GR networks for standard gossip in the synchronous and asynchronous settings are shown in Figs. 4 and 5, respectively. It is assumed that $\alpha = 0.5$ and privacy budget $\epsilon = 1$ without loss of generality. The corresponding privacy lower bounds $\underline{\delta}$ in the x-axis are calculated for the considered ER and GR networks using Theorem 2 given the failure probability f (one-to-one correspondence). Similar results are obtained for private gossip and omitted here due to the space constraint.

Remark 5: Through analysis, it can be seen that the spreading time is inversely proportional to $1-f$ while the privacy lower bound $\underline{\delta}$ is proportional to $1-f$. From Figs. 4 and 5, it can be seen that when $\underline{\delta}$ increases from 0.05 to 0.1, for GR networks, the average spreading time decreases from around 3600 and 2600 to 1800 and 1300 in the synchronous and asynchronous settings, respectively. This means that we can trade a small loss of privacy for dramatic improvement in spreading time. On the other hand, for ER networks or GR networks with large $\underline{\delta}$ (small f), the average spreading time increases slowly as $\underline{\delta}$ decreases. Therefore, the privacy guarantees of gossip protocols can be strengthened with a small loss of spreading time (e.g., the sweet operation points in Figs. 4 and 5), which suggests that methods like adding artificial noise can be useful in privacy-preserving information spreading.

C. Privacy of Gossip Protocols in Delayed Monitoring

In reality, the attacker may not monitor the whole information spreading process right from the beginning. In this section, we try to quantify the differential privacy of general gossip protocols when the monitoring is delayed. To avoid complication, it is assumed that the communications between nodes and the reception at the attacker are perfect. In addition, the attacker knows the global time in the synchronous setting or the number of communication that has occurred in the asynchronous setting since the beginning of information spreading.

Theorem 4: Considering the same setting as in Theorem 1, if the attacker starts monitoring the information spreading process t rounds (or t steps of gossip communications in the asynchronous case) after it begins and $t < D_G$, the gossip-based protocols can guarantee (ϵ, δ)-differential privacy with $\delta \geq \frac{1}{d_{max}^t}\alpha$ in the synchronous setting. In the asynchronous setting

$$\begin{aligned}\delta \geq \max[&\frac{1}{d_{max}^t(t+1)!}\alpha - e^{\epsilon}(1-\alpha)^{D_G-t},\\ &\frac{1}{d_{max}^t(t+1)!}\alpha - e^{\epsilon}\frac{1-\frac{1}{d_{max}^t(t+1)!}\alpha}{n-1}, 0],\end{aligned} \tag{12}$$

in which $d_{max} = max_{i\in V} d_i$ is the largest node degree.

Sketch of proof: In the synchronous setting, consider two nodes i, j such that $d(i,j) = D_G$, and the event that node i's activity is observed by the attacker at the moment when it starts monitoring, which is denoted as $\mathcal{S}_{i,0}$. Considering

another node k such that $d(k,i) = t$, the probability that i is informed at round t is

$$p_G^{(k)}(i \in I_t) \geq \prod_{\substack{m \in p_{k\to i} \\ p_{k\to i}: L(p_{k\to i})=t}} \frac{1}{d_m} \geq (\frac{1}{d_{max}})^t, \qquad (13)$$

where $p_{k\to i}$ is a path from node k to node i and $L(p_{k\to i})$ is the length of this path. Then $p_G^{(k)}(\mathcal{S}_{i,0}) \geq (\frac{1}{d_{max}})^t \alpha$. It is clear that $p_G^{(j)}(\mathcal{S}_{i,0}) = 0$ since it takes at least $D_G > t$ rounds for the information to be delivered to node i from node j. Therefore, by Lemma 1, $\delta \geq (\frac{1}{d_{max}})^t \alpha$.

In the asynchronous setting, again, consider two nodes i, j such $d(i,j) = D_G$, and let $\mathcal{S}_{i,0}$ denote the event that node i's activity is the first one observed by the attacker. If j is the source node, denote the set of informed and active nodes after t steps of communications as $INA_t(j)$. From this set, find the node $k \in INA_t(j)$ that has the shortest path to node i. Clearly, it requires at least $d(k,i)$ $(\geq (D_G - t))$ steps for the information to reach node i from any node in $INA_t(j)$. Consider $O_{INA_t(j)\to i}$ as the event that no communication is observed by the attacker during the process that the information flows from $INA_t(j)$ to node i. Then,

$$\begin{aligned} p_G^{(j)}(\mathcal{S}_{i,0}) &= p_G^{(j)}(\mathcal{S}_{i,0} \bigcap O_{INA_t(j)\to i}) \\ &\leq p_G^{(j)}(O_{INA_t(j)\to i}) \leq (1-\alpha)^{d(k,i)} \leq (1-\alpha)^{D_G - t}. \end{aligned} \qquad (14)$$

Also, considering another node l such that $d(l,i) = t$, the probability that node i is informed at the tth step from the beginning of information spreading is

$$p_G^{(l)}(i \in I_t) \geq (\prod_{\substack{m \in p_{l\to i} \\ p_{l\to i}: L(p_{l\to i})=t}} \frac{1}{d_m}) \frac{1}{t!} \geq \frac{1}{d_{max}^t t!}, \qquad (15)$$

where $\frac{1}{t!}$ is the probability that all nodes in a path $p_{l\to i}$ are activated (whose clocks tick) in a fixed order so that the information reaches node i after t steps from node l. Finally, the probability that node i is activated and its gossip action is observed by the attacker is $\frac{\alpha}{t+1}$. Therefore, $p_G^{(l)}(\mathcal{S}_{i,0}) \geq \frac{\alpha}{d_{max}^t (t+1)!}$. By Lemma 1 and the same logic as Eq. (9), we have Eq. (12). ∎

Remark 6: Gossip protocols are not able to protect the source's identity effectively during the early stage of information spreading. As the spreading process continues, more and more randomness is introduced, leading to stronger and stronger privacy. Therefore, in delayed monitoring, it becomes more difficult for the attacker to identify the source node as the delay increases.

IV. Conclusions and Future Works

In this paper, we investigate the privacy guarantees of gossip-based protocols in general networks. In particular, it is found that source anonymity is closely related to some key network structure parameters, network diameter and decay centrality, in the (arguably) more interesting asynchronous setting. In wireless networks, through a simplified modeling for unreliable communications, the tradeoff between privacy and spreading efficiency is revealed, and it is suggested that natural or artificial interference can enhance the privacy of gossip protocols with the cost of a decrease in spreading speed. Finally, in delayed monitoring, it is verified that the privacy of gossip protocols is enhanced as the delayed time increases, and the corresponding effect is quantified.

Many interesting problems remain open in this line of research besides those already mentioned above. For example, if the attacker is able to measure the distance between any two nodes in the network and rule out those unqualified nodes given existing observations, how will such strategies influence the privacy of gossip protocols? In addition, how can we measure the privacy of gossip protocols against different observation models (e.g., network snapshot [6])? These problems are worth further investigation in future work.

References

[1] R. Chandra, V. Ramasubramanian, and K. Birman, "Anonymous gossip: Improving multicast reliability in mobile ad-hoc networks," in *Proceedings 21st International Conference on DCS*, 2001, pp. 275–283.

[2] A. D. Dimakis, A. D. Sarwate, and M. J. Wainwright, "Geographic gossip: Efficient averaging for sensor networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 3, pp. 1205–1216, 2008.

[3] A. J. Ganesh, A.-M. Kermarrec, and L. Massoulié, "Peer-to-peer membership management for gossip-based protocols," *IEEE transactions on computers*, vol. 52, no. 2, pp. 139–149, 2003.

[4] P. Bianchi and J. Jakubowicz, "Convergence of a multi-agent projected stochastic gradient algorithm for non-convex optimization," *IEEE Transactions on Automatic Control*, vol. 58, no. 2, pp. 391–405, 2013.

[5] Y. Liu, J. Liu, and T. Basar, "Differentially private gossip gradient descent," in *Proceedings of the 57th IEEE CDC*, 2018, pp. 2777–2782.

[6] J. Jiang, S. Wen, S. Yu, Y. Xiang, and W. Zhou, "Identifying propagation sources in networks: State-of-the-art and comparative studies," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 465–481, 2016.

[7] G. Fanti, P. Kairouz, S. Oh, K. Ramchandran, and P. Viswanath, "Hiding the rumor source," *IEEE Transactions on Information Theory*, vol. 63, no. 10, pp. 6679–6713, 2017.

[8] C. Dwork, A. Roth *et al.*, "The algorithmic foundations of differential privacy," *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014.

[9] A. Bellet, R. Guerraoui, and H. Hendrikx, "Who started this rumor? quantifying the natural differential privacy guarantees of gossip protocols," *arXiv preprint arXiv:1902.07138*, 2019.

[10] L. Backstrom, P. Boldi, M. Rosa, J. Ugander, and S. Vigna, "Four degrees of separation," in *Proceedings of the 4th Annual ACM Web Science Conference*, 2012, pp. 33–42.

[11] M. O. Jackson, *Social and economic networks*. Princeton university press, 2010.

[12] H. Zhang, Z. Zhang, and H. Dai, "Gossip-based information spreading in mobile networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 11, pp. 5918–5928, 2013.

[13] M. C. Vuran and I. F. Akyildiz, "Spatial correlation-based collaborative medium access control in wireless sensor networks," *IEEE/ACM Transactions On Networking*, vol. 14, no. 2, pp. 316–329, 2006.

[14] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 2016, pp. 308–318.

[15] J. P. Vilela, M. Bloch, J. Barros, and S. W. McLaughlin, "Wireless secrecy regions with friendly jamming," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 2, pp. 256–266, 2011.

[16] B. Min, S.-H. Gwak, N. Lee, and K.-I. Goh, "Layer-switching cost and optimality in information spreading on multiplex networks," *Scientific reports*, vol. 6, p. 21392, 2016.

[17] A. Picu, T. Spyropoulos, and T. Hossmann, "An analysis of the information spreading delay in heterogeneous mobility dtns," in *2012 IEEE International Symposium on WoWMoM*, 2012, pp. 1–10.