Fundamental Privacy Limits in Bipartite Networks Under Active Attacks

Mahshad Shariatnasab, *Graduate Student Member, IEEE*, Farhad Shirani, *Member, IEEE*, and Elza Erkip, *Fellow, IEEE*

***Abstract*— This work considers active deanonymization of bipartite networks. The scenario arises naturally in evaluating privacy in various applications such as social networks, mobility networks, and medical databases. For instance, in active deanonymization of social networks, an anonymous victim is targeted by an attacker (e.g. the victim visits the attacker's website), and the attacker queries her group memberships (e.g. by querying the browser history) to deanonymize her. In this work, the fundamental limits of privacy, in terms of the minimum number of queries necessary for deanonymization, is investigated. A stochastic model is considered, where 1) the bipartite network of group memberships is generated randomly; 2) the attacker has partial prior knowledge of the group memberships; and 3) it receives noisy responses to its real-time queries. The bipartite network is generated based on linear and sublinear preferential attachment, and the stochastic block model. The victim's identity is chosen randomly based on a distribution modeling the users' risk of being the victim (e.g. probability of visiting the website). An attack algorithm is proposed which builds upon techniques from communication with feedback, and its performance, in terms of expected number of queries, is analyzed. Simulation results are provided to verify the theoretical derivations.**

***Index Terms*— Privacy, social network, bipartite graph, information thresholds, active attack, deanonymization.**

I. Introduction

As TRACKING technologies — both online and in the real-world — become more sophisticated and pervasive, there is a critical need to understand and quantify the resulting privacy risk. For instance, on the internet, users reasonably expect their online identities and web browsing activities to remain private. Unfortunately, this is far from the case in practice; in reality, users are constantly tracked on the internet. Often this is for benign, if somewhat disconcerting, reasons — for instance, websites track users to serve them with targeted digital advertisements [1], [2]. More disturbingly, web tracking can be used to stifle individuals' free speech rights, or target vulnerable minority groups [3]. Furthermore, in wireless applications, the location-based services offered by mobile devices, such as smart phones and autonomous vehicles, can cause significant privacy threats to users, since the time series of locations can be statistically matched to prior user behavior and lead to identification and tracking [4]–[8]. As a result, there is an urgent need to understand and quantify users' privacy risk, that is, what is the likelihood that users can be uniquely identified based on their *fingerprints*? In this work, we study the fundamental limits of privacy in bipartite networks under active attacks. These networks arise naturally in modeling social network group memberships [9]–[11], medical databases [12], and wireless mobility data [5]–[8], among others.

The browser social network deanonymization attack developed by Wondracek *et al.* [9] is a good representative of practical active bipartite network deanonymization (ABND) attacks in the literature, where the attacker runs a malicious website and seeks to deanonymize users who visit the website (see Figure 1). To this end, the attacker first uses a web scraper to scrape the group memberships of users. This serves as the attacker's scanned bipartite graph, $\mathcal{G}_s$, capturing the social network group memberships. Note that the scanned graph might be different from the ground-truth because of users privacy settings that act as a source of noise. When an unknown user (the victim) visits the attacker's website, the attacker queries social network group memberships to find the victim's identity. This is done by using browser history sniffing [13]–[15] to ask questions of the form "is the webpage of social network group 'r_j' in the victim's browser history?" If yes, the attacker assumes that the victim is a member of the social network group r_j, and if no then the attacker assumes the victim is not a member of r_j. Of course, a user might be a member of a group they have not visited, or conversely, might not be a member of a group they have visited; consequently, the attacker's measurement is noisy. The attacker repeats this query for all social network groups in a pre-determined set to obtain the unknown victim's partial fingerprint. By matching the partial fingerprint of query responses to the scanned fingerprints in the scanned graph the victim is deanonymized. In [9], this simple deanonymization strategy is evaluated by using it to find the identities of the users in the Xing social network. It is shown that over 42% of the users who are members of at least one group on Xing (more than 5.7 million users) can be deanonymized successfully

Manuscript received June 8, 2021; revised October 25, 2021; accepted December 21, 2021. Date of publication January 12, 2022; date of current version February 17, 2022. This work was supported in part by NSF under Grant CCF-1815821, Grant CCF-2132843, and Grant CNS-1619129. An earlier version of this paper was presented in part at the Annual Allerton Conference on Communication, Control, and Computing (Allerton) 2017 and in part at the International Symposium on Information Theory (ISIT) 2018. *(Corresponding author: Farhad Shirani.)*

Mahshad Shariatnasab and Farhad Shirani are with the Department of Electrical and Computer Engineering, North Dakota State University, Fargo, ND 58105 USA (e-mail: mahshad.shariatnasab@ndsu.edu; f.shiranichaharsoogh@ndsu.edu).

Elza Erkip is with the Department of Electrical and Computer Engineering, New York University Tandon School of Engineering, Brooklyn, NY 11201 USA (e-mail: elza@nyu.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JSAC.2022.3142299.

Digital Object Identifier 10.1109/JSAC.2022.3142299

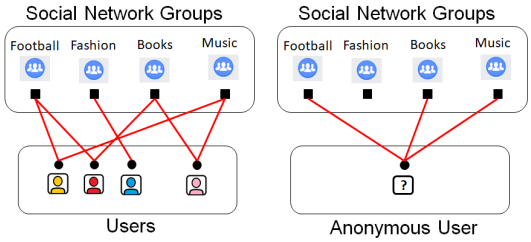


Fig. 1. (Left) An example of a group membership bipartite graph. (Right) An anonymous user (victim) is to be deanonymized based on partial fingerprints.

using the algorithm. Although effective, Wondracek *et al.*'s attack does not answer fundamental questions about the optimal number and type of group memberships to query, and the order in which to issue queries. Other fingerprinting attacks proposed in literature [16]–[20] have also adopted similar ad-hoc approaches without theoretical guarantees or analyses.

A user's fingerprint is the set of group memberships that reflect the user's activities and habits, e.g. websites the user has visited and social network groups that a user is a member of [21], [22], characteristics of the user's web browser (e.g. font size) [23], and physical device features [24]. Fingerprinting based deanonymization attacks build on the empirical observation that, for a large enough set of group memberships, a user's fingerprints are unique. The challenge, from an attacker's standpoint, is that the victim's fingerprints may not be accurately or easily available; i.e., fingerprints may be noisy and the attacker may have to actively query the victim's group memberships, one group at a time, to measure their fingerprint. However, an attacker may only be able to issue a limited number of queries to the victim's device. Our objective is to provide a rigorous mathematical formulation along with theoretical privacy guarantees for the ABND scenario.

In [25], we proposed a mathematical formulation for the ABND problem and introduced a typicality-based strategy by making analogies to the problem of channel coding in information theory, and quantified the amount of information the attacker obtains from each query. We showed that under the assumption that users are equally likely to visit the attacker's website, the total number of queries required for deanonymization grows logarithmically in the number of users. Furthermore, the coefficient of the logarithm is inversely proportional to the mutual information between the random variables corresponding to the scanned graph elements and query responses. In [26], we considered a general distribution, as opposed to a uniform one, on the victim's index among the social network users. This is based on the intuition that more active users would be more likely to visit an attacker's website, resulting in a non-uniform distribution on the victim index. We used techniques from communication over channels with feedback with non-uniform message sets, to propose attack strategies and derive theoretical performance guarantees.

In [25], [26], we considered random bipartite network models in which the edges are independent and identically distributed. However, many bipartite networks of interest, such as social networks [27], [28], networks in cell biology [29], mobility networks [30], and collaboration networks [31], [32] resemble graphs which are generated based on a growing model that grows in accordance to the preferential attachment (PA) rule, first proposed by Simon [33] and rediscovered by Barbási and Albert [34]. In this model, edges are added to the graph iteratively, where at each step, a set of edges are added to the graph randomly such that vertices which have a higher degree are more likely to attract more new connections. In addition to the PA model, another random bipartite graph generation model of interest is the stochastic block (SB) model, where groups are divided into communities, and community memberships of groups affects their likelihood of attracting new users [35], [36]. In this work, we propose a general formulation for the ABND problem, where the bipartite graph random generation model encompasses the PA and SB models, and the scan and query noise models capture the users' different privacy settings and device specifications. We further propose several information-threshold-based deanonymization strategies which build upon the channel coding and hypothesis testing methods studied in [37], [38] to devise deanonymization attacks, and analyze their performance in terms of expected number of queries for successful deanonymization. Our main contributions are summarized below:

- We build upon the ideas in [25], [26] to develop a general mathematical formulation of the ABND problem which encompasses the network generation models such as PA and SB models, and allows for scan and query noises with general distributions. These distributions capture the users' various privacy preferences and device specifications.
- We study the degree distribution and statistical properties of the graph under the proposed generation model. We prove that under certain sparsity conditions on the graph edges — that the number of edges grows linearly in the number of users — the correlation among the user fingerprints is 'weak', so that the fingerprint vector's distribution is well-approximated by a product distribution. These derivations may be of independent interest in the study of bipartite networks.
- We propose information-threshold-based attack strategies and derive theoretical guarantees for theirs success. Roughly speaking, in the proposed strategies, the attacker queries the selected victim's group memberships sequentially and calculates the amount of information obtained, i.e. the amount of uncertainty regarding each user index based on previous query responses. The attack ends when the uncertainty is lower than a given threshold for one of the user indices. The strategy reduces to the one in [26] if the graph edges are assumed to be independent and equally probable, which was proved to be optimal in terms of expected number of queries necessary for successful deanonymization for asymptotically large networks.
- We simulate the performance of the proposed strategies both for synthesized as well as real-world networks, and compare the results with our analytical derivations.

The rest of the paper is organized as follows: Section II describes the notation. In Section III, we provide the problem formulation. In Section IV, we study the degree distribution and other statistical properties of the graph. In Section V, we propose the attack strategy and derive theoretical guarantees for its success. In Section VI, we provide simulation results to verify the theoretical derivations. Section VII, concludes the paper.

II. Notation

We represent random variables by capital letters such as X, U and their realizations by small letters such as x, u. Sets are denoted by calligraphic letters such as $\mathcal{X}$,$\mathcal{U}$. The set of natural numbers, and the real numbers are represented by $\mathbb{N}$, and $\mathbb{R}$ respectively. The random variable $\mathbb{1}_{\mathcal{E}}$ is the indicator function of the event $\mathcal{E}$. The set of numbers $\{n, n+1, \cdots, m\}, n, m \in \mathbb{N}$ is represented by $[n, m]$. Furthermore, for the interval $[1, m]$, we sometimes use the shorthand notation $[m]$ for brevity. For a given $n \in \mathbb{N}$, the n-length vector $(x_1, x_2, \ldots, x_n)$ is written as x^n.

III. Problem Formulation

In this section, we describe our mathematical formulation of the ABND scenario, which generalizes the formulation provided in [25], [26], and encompasses the statistical models for bipartite networks proposed in [39]–[41]. To facilitate explanation, and provide justifications for the model assumptions, we describe the model by focusing on the scenario of deanonymizing social network users using the bipartite network of their group memberships. An ABND attack unfolds in two phases, a passive phase, and an active phase [9], [11], [42]. In the passive phase, the attacker acquires a noisy observation of the bipartite network of group memberships by scanning the whole social network. In the active phase, the attacker targets a specific victim (e.g. a user visiting the attacker's website), and uses browser history sniffing techniques to query the victim's group memberships. The attacker constructs a *fingerprint* for the victim using the (noisy) query responses, and identifies the victim by comparing this fingerprint with the noisy scan of the bipartite graph acquired in the passive phase of the attack. As shown in Figure 2, the model consists of three components which are described in detail in the following sections: i) the *ground-truth* $\mathcal{G}_0$ representing the 'true' group memberships of users in the social network (Section III-A), ii) the *scanned graph* $\mathcal{G}_s$ which represents the attacker's prior knowledge of the ground-truth (Section III-B), and iii) the *query responses*, represented by $\mathcal{G}_q$, which are acquired by the attacker by querying the victim in the active phase of the attack (Section III-C). The objective is to design an attack strategy which determines the sequence of queries made by the attacker to deanonymize the victim, along with theoretical guarantees for its success (Section III-D).

A. The Ground-Truth

The collective set of group memberships in the social network are called the *ground-truth*. The ground-truth is represented by a bipartite graph.

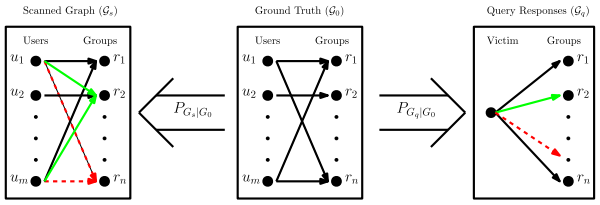


Fig. 2. Components of the ABND problem: i) the ground-truth characterized by the bipartite graph $\mathcal{G}_0$ and generated based on $P_{\mathcal{G}_0}$, ii) the scanned graph $\mathcal{G}_s$ generated based on $P_{\mathcal{G}_s|\mathcal{G}_0}$, and iii) the query responses $\mathcal{G}_q$ generated based on $P_{\mathcal{G}_q|\mathcal{G}_0}$. The black edges represent 'true' group memberships, whereas green and dashed-red edges show additions and omissions, respectively, which may manifest due to noise in scanning the social network in passive phase of the attack, and noisy query responses in the active phase.

Definition 1 (Bipartite Graph): A bipartite graph $\mathcal{G} = (\mathcal{V}_1, \mathcal{V}_2, \mathcal{E})$, is a graph with vertex set $\mathcal{V}_1 \bigcup \mathcal{V}_2$ and edge set $\mathcal{E} \subseteq \{(v_i, v_j) | v_i \in \mathcal{V}_1, v_j \in \mathcal{V}_2\}$, where $\mathcal{V}_1 \cap \mathcal{V}_2 = \phi$.

We consider a social network with user set $\mathcal{U} \triangleq \{u_1, u_2, \cdots, u_m\}, m \in \mathbb{N}$, and group set $\mathcal{R} \triangleq \{r_1, r_2, \cdots, r_n\}, n \in \mathbb{N}$. The ground-truth is characterized by a bipartite graph $\mathcal{G}_0 = (\mathcal{U}, \mathcal{R}, \mathcal{E})$, where ($\mathcal{U}$,$\mathcal{R}$) partitions the vertex set, and the edge set $\mathcal{E}$ consists of all pairs $(u_k, r_j), k \in [m], j \in [n]$ for which user u_k is a *member* of the group r_j.

Definition 2 (Group Size): Let the set of users which are members of the jth group $r_j, j \in [n]$ be denoted by $\mathcal{U}_j \triangleq \{u_{k_1}, u_{k_2}, \cdots, u_{k_{D_j}}\}, k_1, k_2, \cdots, k_{D_j} \in [m]$. Then, $D_j \triangleq |\mathcal{U}_j|$ is called the size of group r_j.

Example 1: In the Facebook social network, $\mathcal{U}$ is the set of users and $\mathcal{R}$ includes the pages/ events/ groups/ applications on Facebook. Here, the groups under consideration are those whose member lists are publicly available.

Each user is assigned a fingerprint based on its group memberships. The fingerprint is a binary vector of indicator functions, indicating the membership of the user in each particular group. Alternatively, the user's fingerprint is the vector of indicator functions corresponding to the edges between the user and each of the groups.

Definition 3 (Fingerprint): Consider the ground-truth bipartite graph $\mathcal{G}_0 = (\mathcal{U}, \mathcal{R}, \mathcal{E})$:

- *For a user $u_k, k \in [m]$, the set $\mathcal{R}_k \triangleq \{r_j | (u_k, r_j) \in \mathcal{E}\}, k \in [m]$ is called the set of groups associated with u_k.*
- *The fingerprint of user $u_k, k \in [m]$ is the vector $(R_{k,j})_{j \in [n]} \triangleq (R_{k,1}, R_{k,2}, \cdots, R_{k,n})$, where*

$$R_{k,j} \triangleq \begin{cases} 1 & \text{if } r_j \in \mathcal{R}_k \\ 0 & \text{otherwise} \end{cases}, \qquad k \in [m],\ j \in [n].$$

- *The vector $R_{k,\mathcal{I}} \triangleq (R_{k,j})_{j \in \mathcal{I}}$ is called a partial fingerprint of $u_k, k \in [m]$, where $\mathcal{I} \subseteq [n]$.*

We consider a stochastic model which is a generalization of those considered in prior works on active social network deanonymization [11], [25], [26], and includes as a special case several statistical models such as SB model, and PA model which have been used for bipartite networks such as social network group memberships, collaboration networks, authorship networks, and location networks [39]–[41].

The ground-truth $\mathcal{G}_0$ is generated iteratively based on a 'growing network' model as follows. Fix $\mu \in \mathbb{N}$, and define $\Delta \triangleq \mu n$, where n is the number of social network groups. The iterative process is initiated by considering a bipartite graph $(\mathcal{U}, \mathcal{R}, \phi)$, which has no edges connecting its two sets of vertices. The ground-truth graph is generated in Δ iterative steps, where at each step a single edge is added to the graph, so that $|\mathcal{E}| = \Delta$ after the last iteration. As a result, the average group size is equal to $\frac{\Delta}{n} = \mu$. For $t \in [\Delta]$, define $\mathcal{G}_0(t) \triangleq (\mathcal{U}, \mathcal{R}, \mathcal{E}(t))$ as the bipartite graph at step t. The group membership sets at step $t \in [\Delta]$ are denoted by $\mathcal{U}_j(t), j \in [n]$, and the group sizes are denoted by $D_{t,j} \triangleq |\mathcal{U}_j(t)|$. Building upon the idea of PA graph generation models — where the likelihood that a given vertex connects to a new vertex is linearly related with the degree of that vertex — we assume that, at each step, groups attract new members in accordance with their *popularity* at that step. To elaborate, we assume that each group $r_j, j \in [n]$ is assigned a popularity value $\tau_j(t)$ which captures its popularity at time t. The value of $\tau_j(t)$, which may depend on the size of group r_j among other factors, affects the probability of r_j attracting new members as described in the sequel. In this work, we restrict to the case where the value of $\tau_j(t), j \in [n], t \in [\Delta]$ depends only on the group size $D_{t,j}$ and an initial value $\tau_j(0)$. The vector $\underline{\tau}(t) = (\tau_1(t), \tau_2(t), \cdots, \tau_n(t))$ represents the vector of group popularity values at time t.

Initiation: Each group $r_j, j \in [n]$ is assigned an initial popularity value $\tau_j(0) > 0$. The ground-truth graph is initiated as $\mathcal{G}_0(0) \triangleq (\mathcal{U}, \mathcal{R}, \phi)$. So, the group membership sets are $\mathcal{U}_j(0) = \phi, j \in [n]$ and $D_{0,j} = 0, j \in [n]$.

Step t: At each step $t \in [\Delta]$, a group r_{J_t} and a user u_{K_t} are chosen as described next, and the corresponding edge (u_{K_t}, r_{J_t}) is added to the bipartite graph, i.e. $\mathcal{E}(t) = \mathcal{E}(t-1) \cup \{(u_{K_t}, r_{J_t})\}$. First, a group r_{J_t} is chosen among the set of all groups $\mathcal{R}$ according to the probability distribution $\mathbf{P}(t) = (P_1(t), P_2(t), \cdots, P_n(t))$ defined below:

$$P_j(t) \triangleq \frac{\tau_j(t-1)}{\sum_{j'=1}^{n} \tau_{j'}(t-1)},$$

Next, a user u_{K_t} is chosen randomly and uniformly from the set of users which are not members of r_{J_t}, i.e. $[m] - \mathcal{U}_{J_t}(t-1)$. The edge (u_{K_t}, r_{J_t}) is added to the edge set. The group popularity values are updated as follows:

$$\tau_j(t) = \begin{cases} \tau_j(t-1) & \text{if } j \neq J_t \\ f(\tau_j(t-1), \tau_j(0)) & \text{if } j = J_t, \end{cases} \quad j \in [n] \qquad (1)$$

where $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is a strictly increasing function which captures the increase in a group's popularity due to the addition of a new member and its subsequent effect on the group's attractiveness to new members. For tractability, we assume that $f(\cdot, \cdot)$ is the same for all groups and fixed over time. If $f(x, y), x, y \in \mathbb{R}$ is a linear function of x for any fixed y, then we recover the PA model in [33], [34]. On the other hand, if $f(x, y)$ is concave in x for any fixed y, then an increase in the popularity of an unpopular group increases its attractiveness to new users more significantly than a similar increase in the popularity of an already popular group. On the other hand, a convex $f(\cdot, \cdot)$ creates the opposite effect.

Remark 1: We have assumed that at each step, there exists a user which is not already a member of r_{J_t}. We will show that due to the sparsity conditions considered in this work, the probability that there exists a group for which every user is its member, vanishes exponentially in the number of users as the graph becomes larger (Proposition 2). However, for completeness, we assume that if every user is already a member of r_{J_t} (i.e. if $\mathcal{U}_{J_t}(t-1) = [m]$), then an edge is not added in this step, the group popularities are updated as usual, and the generation process advances to the next step.

Remark 2: We study bipartite graphs where the edges are binary-valued, i.e. a single edge between a given user and a given group is either present or absent. A natural extension is to consider edges with non-binary attributes and multi-graphs. The attribute captures the nature of a users' group membership, e.g. group administrator, active member, etc. Inclusion of such information in the network graph may assist the attacker in deanonymizing the victim. The information theoretic derivations provided in the next sections can be extended in a straightforward manner to graphs with attributed edges and multigraphs, where attributes are taken from an arbitrary finite set, and a finite number of edges is allowed between each two vertices, respectively.

In this work, we focus on the particular choice of $f(x, y) = ((x-y)^{\frac{1}{\alpha}} + 1)^{\alpha} + y, \alpha \in (0, 1]$. This choice recovers several models for bipartite networks studied in prior works —- such as equiprobable edges model, SB model, and linear and sublinear PA model — by taking different values of α as described next. The parameter α is an intrinsic network parameter. In this case, Equation (1) can be rewritten as:

$$\tau_j(t) = \begin{cases} \tau_j(t-1) & \text{if } j \neq J_t \\ D_{t-1,j}^{\alpha} + \tau_j(0) & \text{if } j = J_t, D_{t-1,j} < n \\ \tau_j^{\alpha}(t-1) + \tau_j(0) & \text{otherwise,} \end{cases}$$

where $j \in [n]$, and $t \in [\Delta]$. At a high level, α determines the effect of the groups' sizes on the membership choices of new users, where larger α means that the group-size plays a significant role in attracting new users, with large groups being more attractive, and at the other end of the spectrum, if $\alpha \to 0$, then the group popularities are constant through the generation process regardless of the group sizes. We focus on $\alpha \leq 1$ which leads to linear or sublinear PA and has been shown to be a suitable model for various networks of interest [27]–[32].

Definition 4 (Ground-Truth Parameters): The ground-truth statistics are parametrized by $(n, m, \alpha, \Delta, (\tau_j(0))_{j \in [n]})$. The following scenarios are considered in this work:

α-Preferential Attachment (α-PA): *This is a generalization of the PA model, where $f(x) = ((x-y)^{\frac{1}{\alpha}} + 1)^{\alpha} + y, \alpha \in (0, 1]$ and initial popularities are $\tau_j(0) = \tau_{j'}(0) = 1, j, j' \in [n]$.*

Stochastic Blocks (SB): *We take $\alpha \to 0$ and $\tau_j(0) \in \mathcal{T}$, where $\mathcal{T}$ is a finite set. The collection of subsets $\mathcal{C}_\tau = \{r_j : \tau_j(0) = \tau\}, \tau \in \mathcal{T}$ are called the communities of social network groups.*

Remark 3: As a special case of the SB model, let us take $\alpha \to 0$ and $\tau_j(0) = \tau_{j'}(0), j, j' \in [n]$. Then, $f(x, y) = x$ for all $x, y \in \mathbb{R}$, and $\tau_j(t) = \tau_{j'}(t), j, j' \in [n], t \in [\Delta]$.

We call this the Independent and Equiprobable Edges (IEE) scenario. This is analogous to the Erdös-Rényi model for non-bipartite graphs [43], and was studied in [26]. In this case, $P_j(t) = P_{j'}(t), j \in [n], t \in [\Delta]$, and the groups are equally likely to attract new users regardless of their current number of members.

Remark 4: In the SB scenario, we have $P_j(t) = \frac{\tau}{\sum_{\tau' \in \mathcal{T}} \tau' |\mathcal{C}_{\tau'}|}, r_j \in \mathcal{C}_\tau, t \in [\Delta], \tau \in \mathcal{T}$. So, the groups which belong to the same community $\mathcal{C}_\tau, \tau \in \mathcal{T}$ are equally likely to attract new users regardless of their current number of members. Groups may be classified into different communities based on the shared interests of their users, e.g. age group, profession, etc. This model resembles the stochastic block model for social network friendship graphs [35], [36].

Remark 5: In the α-PA scenario, if $\alpha = 1$, we have $f(x, y) = x - y + 1, x, y \in \mathbb{R}$ and the model becomes the well-studied (linear) PA model. In this case, with appropriate choice of initial popularities, the group sizes follow a power-law. This is in agreement with empirical studies of social network group memberships (e.g. [27], [28]), where such power-law behavior has been observed.

Remark 6: In practice, the ground-truth statistics parametrized by $(n, m, \Delta, (\tau_j(0))_{j \in [n]}, \alpha)$ are not available to the attacker. Rather, the attacker acquires an estimate of these parameters as in [39] based on prior observations of the bipartite network.

B. The Scanned Graph

As described in previous sections, the first phase of the fingerprinting attack is the passive phase, in which the attacker scans the social network for publicly available information regarding the users' group memberships. The attacker's observation of the ground-truth, acquired through this scanning process, is represented by the bipartite graph $\mathcal{G}_s = (\mathcal{U}, \mathcal{R}, \mathcal{E}_s)$, which is a partial and noisy observation of the users' group memberships. One reason for the noise in the scanned graph is that some users may have made a subset of their group memberships hidden which results in edge omissions in the scanned graph. We model the resulting noise stochastically by assuming that the set of edges $\mathcal{E}_s$ in the scanned graph is generated randomly, conditioned on the set of edges $\mathcal{E}_0$ in the ground-truth graph. As discussed above, the difference between $\mathcal{E}_0$ and $\mathcal{E}_s$ is due to the privacy preferences of a specific user. As a result, we assume that the noise statistics in scanning a specific user-group edge (u_k, r_j) is dependent on the corresponding user preference which is captured by the parameter $\gamma(k) \in \Gamma$, where Γ is a finite set. This is formalized below.

Definition 5 (Scanned Graph Statistics): Let $P^{\gamma(k)}_{E_s|E_0}(\cdot|\cdot), \gamma(k) \in \Gamma, k \in [m]$ be a collection of conditional probability distributions, where E_s and E_0 take binary values, and Γ is a finite set. Let $R_{k,j} \triangleq \mathbb{1}((u_k, r_j) \in \mathcal{E}_0)$ and $F_{k,j} \triangleq \mathbb{1}((u_k, r_j) \in \mathcal{E}_s), k \in [m], j \in [n]$. Then,

$$P(\mathcal{E}_s|\mathcal{E}_0) = \prod_{k \in [m], j \in [n]} P^{\gamma(k)}_{E_S|E_0}(F_{k,j}|R_{k,j}).$$

In particular, the following Markov chains are assumed:

$$F_{k,j} \leftrightarrow R_{k,j}, k \leftrightarrow (F_{k',j'}, R_{k',j'})_{(k',j') \neq (k,j)}, \quad k \in [m], \quad j \in [n].$$

Example 2 (Erasure Model for $\mathcal{G}_s$): Assume that the attacker scans a social network to acquire the scanned graph. The attacker observes a subset of the true group memberships of users [9] since some users choose to keep their membership in certain groups private. As a result, the scanned graph $\mathcal{G}_s$ consists of a sampled subset of the edges in the ground-truth $\mathcal{G}_0$. For simplicity, let us assume that the membership of user u_k in group r_j is publicly available with probability $1 - s_k, k \in [m], j \in [n]$, where $s_k \in [0, 1]$. Then,

$$Pr(\mathcal{E}_s|\mathcal{E}_0) = \mathbb{1}(\mathcal{E}_s \subset \mathcal{E}_0) \times \prod_{k \in [m]} s_k^{|\mathcal{R}'_k|}(1 - s_k)^{|\mathcal{R}_k| - |\mathcal{R}'_k|},$$

where $\mathcal{R}_k$ and $\mathcal{R}'_k$ are the groups in which $u_k, k \in [n]$ is a member of in $\mathcal{G}_0$ and $\mathcal{G}_s$, respectively.

Remark 7: We assume that the attacker does not have knowledge of the users' privacy preferences, i.e., it does not know the value of $\gamma(k), k \in [m]$ in Γ. The attacker only has access to the statistics $P^{\gamma}_{E_s|E_0}, \gamma \in \Gamma$.

C. Query Responses

In the active phase of the attack, the attacker targets a victim, and actively queries its group memberships. For instance, the victim visits a malicious website, and the attacker uses browser history sniffing techniques to query the victim's group memberships. The attacker may query the victim's group memberships sequentially by sending a single query regarding the victim's membership in a group at each step of the active attack, receiving a response, and deciding on the next query [9]. Alternatively, it may query a batch of group memberships simultaneously [13]–[15]. In this work, we focus on the first scenario, where the queries are made sequentially, one after the other. However, the analysis can be extended to the second scenario, where queries are made in batches, in a straightforward manner.

The objective is to deanonymize the victim based on their group membership fingerprint. We model the victim stochastically by assuming that it is chosen randomly from the user set. In general, the users are not equally likely to be a victim of an attack, For instance, users are not equally likely to visit a malicious website, risk-averse users are less likely to be the victim of a fingerprinting attack compared to risk-taker users. As a result, we assume that the victim u_M is chosen from $\mathcal{U}$ based on an underlying distribution P_M.

Remark 8: In this work, following the conventional approach in privacy and security literature, we investigate a 'genie-aided' attacker by assuming access to P_M in order to derive theoretical guarantees for users' privacy. However, it should be noted that, in practice, the attacker may only have an estimate $\hat{P}_M$ of P_M or it may not have any prior knowledge of these statistics at all. In such cases, the attack strategies investigated in the following sections may be extended naturally, and their probability of success can evaluated with respect to a 'worst-case' distribution $\hat{P}_M$.

Let us assume that the attacker queries the group memberships of the victim u_M in the sequence of groups $(r_{j_1}, r_{j_2}, \cdots, r_{j_\ell}), j \in [n]$ in $\ell \in \mathbb{N}$ queries, and receives the binary vector of query responses $Y_1, Y_2, \cdots, Y_\ell$, where $Y_i = 1$ indicates a positive response and $Y_i = 0$ a negative response. Generally, query responses are noisy since browser history sniffing techniques are imperfect and only provide noisy observations of the victim's browsing history. That is, Y^ℓ is a noisy version of the true group membership indicators $(R_{j_1}, R_{j_2}, \cdots, R_{j_\ell})$. The noise statistics are determined by the users' software (e.g. browser [14]) and hardware specifications (e.g. CPU and memory specifications [13]), and depend on the type of history sniffing attack. However, these statistics do not depend on the specific website or group whose membership is being queried. This dependency is captured by the parameter $\theta(M)$, where $\theta : [m] \to \Theta$, and Θ is a finite set. The following definition formalizes the stochastic model for the query responses.

Definition 6 (Noisy Query Responses): Let $\ell \in \mathbb{N}$ and let $P^\theta_{Y|R}, \theta \in \Theta$ be a collection of probability distributions, where Y and R are binary variables and Θ is a finite set. For the sequence $j_1, j_2, \cdots, j_\ell \in [n]$, assume that victim's fingerprint is $(R_{j_1}, R_{j_2}, \cdots, R_{j_\ell})$ and the received query responses are $Y_1, Y_2, \cdots, Y_\ell$. Then,

$$P(Y^\ell = y^\ell | (R_{j_i})_{i\in[\ell]} = r^\ell) = \prod_{i=1}^{\ell} P^{\theta(M)}_{Y|R}(y_i|r_i), \qquad y^\ell, r^\ell \in \{0,1\}^\ell,$$

where the parameter $\theta(M)$ takes values from Θ and its value depends on the victim's index M.

Remark 9: In practice, the attacker does not have access to the statistics $P^{\theta(k)}_{Y|R}(y_i|r_i), k \in [m]$. Rather, it may query the victim's software and hardware specifications to acquire $\theta(k)$, and then estimate the noise statistics based on prior observations of the querying process with these specifications and based on the history sniffing technique used by the attacker. This is in contrast with the noise model in the scanned graph $P^{\gamma(k)}_{Y|R}(y_i|r_i), k \in [m]$, where the attacker has no means of learning the user's privacy preferences $\gamma(k)$.

To summarize, an active bipartite network deanonymization setup is characterized as follows.

Definition 7 (Active Bipartite Network Deanonymization): An active bipartite network deanonymization setup is characterized by parameters $(n, m, \Delta, \Theta, \Gamma, \alpha, (\tau_j(0))_{j\in[n]}, P_M, (P^{\gamma(k)}_{E_S|E_0})_{k\in[m],\gamma\in\Gamma}, (P^{\theta(k)}_{Y|R})_{k\in[m],\theta\in\Theta})$, where n is the number of groups, m the number of users, P_M determines the victim's (u_M) distribution among the users $\mathcal{U}$, $P^{\theta(k)}_{Y|R}, \theta(k) \in \Theta$ is the query response noise statistics for user $k \in [m]$, $P^{\gamma(k)}_{E_S|E_0}, \gamma(k) \in \Gamma$ is the scanned graph noise statistics for user $k \in [m]$, α is the network growth parameter, Δ is the total number of edges, and $(\tau_j(0))_{j\in[n]}$ are the initial group popularities.

D. Attack Strategy

Given the scanned graph $\mathcal{G}_s$ acquired by scanning the ground-truth $\mathcal{G}_0$, the attacker's objective is to identify the victim using the minimum number of queries possible, and with small probability of error. An attack strategy determines the sequence of queries made by the attacker, and identifies the victim based on the query responses. It consists of a sequence of query functions $x_t(\cdot,\cdot), t \in \mathbb{N}$ and identification functions $Id_t(\cdot,\cdot), t \in \mathbb{N}$, where at time[1] t, the query function $x_t(\mathcal{G}_s, Y^{t-1})$ takes the scanned graph $\mathcal{G}_s$ and the received query responses Y^{t-1} as input, and outputs the group $r_{j_t} = x_t(\mathcal{G}_s, Y^{t-1}), j_t \in [n]$ whose connection with the victim is to be queried next. Assume that the response Y_t is received. The identification function $Id_t(\mathcal{G}_s, Y^t)$ compares the received query responses Y^t with the users' fingerprints in the scanned graph $\mathcal{G}_s$, and either outputs the identity of the victim, or indicates that the identity cannot be determined yet, hence the attack continues with the next query. This is formalized below.

Definition 8 (Attack Strategy): Consider an ABND scenario parametrized by $(n, m, \Delta, \Theta, \Gamma, \alpha, (\tau_j(0))_{j\in[n]}, P_M, (P^{\gamma(k)}_{E_S|E_0})_{k\in[m],\gamma\in\Gamma}, (P^{\theta(k)}_{Y|R})_{k\in[m],\theta\in\Theta})$. An attack strategy consists of a sequence of query functions $x_t : \{0,1\}^{m\times n} \times \{0,1\}^{(t-1)} \to \mathcal{R}, t \in \mathbb{N}$ and identification functions $Id_t : \{0,1\}^{m\times n} \times \{0,1\}^t \to \mathcal{U} \cup \{e\}$, where $x_t(\mathcal{G}_s, Y^{t-1})$ outputs the group whose edge connection with the victim is queried at time t, and $Id_t(\mathcal{G}_s, Y^t)$ either outputs the victim's identity among the user set $\mathcal{U}$ or outputs 'e' in which case further queries are made and the attack continues. Let $Q = min\{t \in \mathbb{N} : Id_t(\mathcal{G}_s, Y^t) \in \mathcal{U}\}$. Then, the probability of error P_e and expected number of queries $\overline{Q}$ are defined as:

$$P_e((x_t, Id_t)_{t\in\mathbb{N}}) \triangleq P(Id_Q(\mathcal{G}_s, Y^Q) \neq u_M)$$
$$\overline{Q}((x_t, Id_t)_{t\in\mathbb{N}}) \triangleq \mathbb{E}(Q),$$

where the probabilities are with respect to $M, \mathcal{G}_0, \mathcal{G}_s$ and $Y_t, t \in [Q]$.

Definition 9 (Minimum Expected Queries): For the ABND problem characterized by $(n, m, \Delta, \Theta, \Gamma, \alpha, (\tau_j(0))_{j\in[n]}, P_M, (P^{\gamma(k)}_{E_S|E_0})_{k\in[m],\gamma\in\Gamma}, (P^{\theta(k)}_{Y|R})_{k\in[m],\theta\in\Theta})$, and error probability $\epsilon > 0$, the minimum expected number of queries is defined as:

$$Q^*_\epsilon \triangleq \inf_{(x_t, Id_t)_{t\in\mathbb{N}}} \{\overline{Q}((x_t, Id_t)_{t\in\mathbb{N}}) | P_e((x_t, Id_t)_{t\in\mathbb{N}}) \leq \epsilon\}.$$

Our objective is to investigate the necessary and sufficient conditions under which an attacker can deanonymize the victim reliably (i.e. with vanishing error probability) over asymptotically large bipartite networks. That is, we want to investigate the problem when the number of users m grow asymptotically large. In particular, based on observations of real-world social networks (e.g. [39], [44]), we investigate the ABND problem under the following asymptotic regime:

- **Number of Groups:** The number of groups n grows linearly in m, i.e. $m = \beta n$ for a fixed $\beta > 0$.
- **Noise Parameters:** The sets Θ, Γ and $P^\theta_{Y|R}, P^\gamma_{E_s|E_0}, \theta \in \Theta, \gamma \in \Gamma$ are fixed in m. This is justified since $P^\gamma_{E_s|E_0}$

[1]Note that we have used the variable 't' to refer to two different time quantities. One is the steps in the ground-truth generation process ($t \in [\Delta]$) in Section III-A, and the other one is the number of queries sent in the active phase of the attack ($t \in \mathbb{N}$) which is discussed here.

and $P^{\theta}_{Y|R}$ are determined by the users' privacy preference options in the social network, and their software/hardware specifications, respectively, and do not change as the number of users increases asymptotically.

- **Sparsity:** The average number of groups in which any given user is a member of is constant as the network grows. That is, $\Delta = \mu n = \frac{\mu}{\beta} m, \mu \geq 1$, so that the average group size μ is constant in n.
- **Victim's Distribution:** The users' likelihood of being the victim decreases inversely in m, that is $P_M(u_k) = \frac{c_k}{m}$, where $\sum_{k\in[m]} c_k = m$ and $c_k < \lambda, k \in [m]$ as $n \to \infty$ for some constant $\lambda > 0$.

Remark 10: As mentioned in the prequel, the deanonymization scenario considered in this work is closely related to problems in hypothesis testing and communication over channels with feedback (e.g. [37], [38]), and some of the proof techniques presented here build upon those prior works. An additional application scenario which is related to this work is the guesswork problem, where the attacker's objective is to guess the victim's fingerprint vector rather than their identity (e.g., [45]). Exploring the connections between these two problems is an interesting avenue of future research.

IV. Memory Structure of the Ground-Truth Edges

The scanned graph and the query responses provide the attacker with two noisy observations of the victim's group membership fingerprint in the ground-truth. The attacker identifies the victim by reconciling the query responses with the user fingerprints in the scanned graph and finding a unique match (e.g. jointly typical pair of fingerprint and query response vectors). One major obstacle in analyzing the fundamental performance limits of attack strategies is the *memory structure* in the user's fingerprint induced due to the generation model of the ground-truth described in Section III-A. That is, the generation model induces correlation among the users' membership in different groups. This prohibits the conventional methods such as type analysis and large deviations techniques which have been used in deriving theoretical performance limits in similar scenarios in group testing [38] and communications [37] problems, as well as the analysis techniques in prior work on ABND [25], [26]. In this section, we show that under the sparsity assumption on the total number of edges that $\Delta = \mu n$, the memory in the users' fingerprint is weak, so that its joint distribution is well-approximated by a product distribution. The derivations are used in the next sections, where we propose attack strategies and derive sufficient conditions for their success. These are also of independent interest in analyzing degree distributions of vertices in bipartite networks.

A. Weakly Correlated Group Sizes

Let us recall that the size of group $r_j, j \in [n]$ at step $t \in [\Delta]$ of the generation process is defined as $D_{t,j} = |\mathcal{U}_j|, j \in [n]$. As a first step towards investigating the correlation among users' memberships in different groups, we study the joint moments of $(D_{\Delta,j})_{j\in[n]}$ and show that they converge to a finite constant as $n \to \infty$, and $\Delta = \mu n \to \infty$.

Proposition 1 (Group Size Correlation): Let $0 < \alpha < 1$. For a ground-truth graph generated according to the α-PA model, the following holds:

$$\mathbb{E}(D_{\Delta,j}) = \mu, \quad j \in [n], \tag{2}$$

$$\mathbb{E}(D^2_{\Delta,j}) = O(1), \quad j \in [n], \tag{3}$$

$$\mathbb{E}(D_{\Delta,i} D_{\Delta,j}) = \mu^2 + O(\frac{1}{n}), \quad i \neq j, \tag{4}$$

$$\mathbb{E}(D_{\Delta,1} D_{\Delta,2} \cdots D_{\Delta,\zeta}) = \mu^{\zeta}(1 + \zeta O(\frac{1}{n})), \quad \zeta \in [n], \tag{5}$$

$$\mathbb{E}(D^2_{\Delta,1} D_{\Delta,2} D_{\Delta,3} \cdots D_{\Delta,\zeta}) \leq \mu^{\zeta-1} \mathbb{E}(D^2_{1,\Delta}), \quad \zeta \in [n], \tag{6}$$

$$\mathbb{E}(D_{\Delta,1} D_{\Delta,2} D_{\Delta,3} \cdots D_{\Delta,\zeta}) \leq \mu^{\zeta}, \quad \zeta \in [n]. \tag{7}$$

Proof: Please see [46]. □

Proposition 1 can be interpreted as follows: Equation (2) shows that $\mathbb{E}(D_{\Delta,j}), j \in [n]$, the average size of the jth group after Δ generation steps, is equal to $\mu = \frac{\Delta}{n}$. This holds due to symmetry in the ground-truth generation process. Equation (3) shows that the variance of the jth group size $\mathbb{E}(D^2_{\Delta,j}) - \mathbb{E}^2(D_{\Delta,j}), j \in [n]$ is bounded from above as the number of groups is increased asymptotically. Equation (4) shows that the group sizes have weak pairwise correlation. That is, they become uncorrelated as the number of groups is increased asymptotically. This follows from the sparsity assumption, $\Delta = \mu n$, described in Section III. Equations (5) and (7) generalize Equation (4) and show that the sizes of any finite subset of groups have weak joint correlation and become uncorrelated as the number of groups grows asymptotically. Equation (6) generalizes Equation (5) to some of the higher joint moments of the group sizes, and is used in the derivations in the sequel.

B. Almost Memoryless Fingerprints

Next, we prove that under the sparsity condition $\Delta = \mu n$, the fingerprints in the ground-truth are 'almost' memoryless. Let the number of groups in which a user is a member be denoted by $C_i \triangleq |\mathcal{R}_i|, i \in [m]$. The users' memberships in different groups are correlated due to the ground-truth generation model. We are interested in investigating this correlation. As a first step, we show in the following that each user's fingerprint is sparse (i.e. has few ones).

Proposition 2 (Sparsity of the User Fingerprint Vector): Let $\alpha \in (0, 1]$, $\mu \in \mathbb{N}$, and $\beta > 0$. For a ground-truth graph generated according to the α-PA model with $n \in \mathbb{N}$ groups, $m = \beta n$ users, and $\Delta = \mu n$ edges, there exists a constant $c > 0$ such that:

$$P(C_i \geq \ell) \leq c 2^{-n D_b(\frac{\mu}{m}(1+\psi)||\frac{\mu}{m})}, \tag{8}$$

where $\ell = \frac{1}{\beta}\mu(1+\psi)$, $\psi \in (0, \frac{m}{\mu} - 1)$, and $D_b(p||q) = p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}$ is the binary Kullback-Leibler divergence. In particular, let $\psi_n > 0, n \in \mathbb{N}$ such that $\psi_n = \omega(1)$. Then,

$$P(C_i \geq \psi_n) \to 0, \quad \text{as } n \to \infty. \tag{9}$$

Proof: The proof follows by using an extension of Hoeffding's inequality for weakly correlated variables given in [47], along with (7) which shows that the group sizes are weakly jointly correlated, i.e. their joint correlation is small. The complete proof is provided in [46]. □

The next proposition shows that the distribution of the fingerprint of each user in the ground-truth graph is close to a memoryless distribution.

Proposition 3 (Memoryless Fingerprints in α-PA): Let $\alpha \in (0,1]$. For a ground-truth graph generated according to the α-PA model, consider the partial fingerprint $\mathbf{R} \triangleq (R_{i,j_k})_{k\in[n']}, j_k \in [n], n' \in [n]$ of user $u_i, i \in [m]$. The following holds:

$$(1-\frac{n'\mu}{m})\prod_{k=1}^{n'} P_R(s_k) \le P_{\mathbf{R}}(s^{n'}) \le e^{\frac{\mu}{\beta}} \prod_{k=1}^{n'} P_R(s_k), \quad s^{n'} \in \{0,1\}^{n'},$$

where $P_R(\cdot) = P_{R_{i,j}}(\cdot), i \in [m], j \in [n]$. Furthermore, assume that $n' > \frac{m}{\mu}$ and $\sum_{i=1}^{n'} \mathbb{1}(s_i = 1) = o(n)$ for some constant finite number $C > 0$. Then, there exists $c' > 0$ whose value only depends on μ and β such that:

$$c' \prod_{k=1}^{n'} P_R(s_k)(1+o(1)) \le P_{\mathbf{R}}(s^{n'}) \le \prod_{k=1}^{n'} P_R(s_k)(1+o(1)), \quad s^{n'} \in \{0,1\}^{n'},$$

as $n \to \infty$.

Proof: The proof relies on the fact that due to the generation process described in Section III, given that the size of the jth group is $D_j = d_j, j \in [n], d_j \in [m]$, the probability that the user $u_k, k \in [m]$ is in group r_j is $\frac{d_j}{m}$ independent of all other users' memberships, i.e. $P((u_k, r_j) \in \mathcal{E}_0 | D_j = d_j) = \frac{d_j}{m}$. The complete proof is provided in [46]. □

In the SB scenario, edge probabilities do not change during the generation process and the number of groups associated with each user follows a (truncated) Binomial distribution with parameters $(\Delta, \frac{\mu}{\Delta})$. As a result, it is straightforward to establish the memoryless property of the fingerprints using standard arguments based on law of large numbers. It should be noted that there is correlation among group sizes in this case since for instance

$$\mathbb{E}(D_{1,\Delta}D_{2,\Delta}) = \mathbb{E}(D_{1,\Delta}\mathbb{E}(D_{2,\Delta}|D_{1,\Delta})) = \mathbb{E}(D_{1,\Delta})\mathbb{E}(D_{2,\Delta})(1-\frac{\mathbb{E}(D_{1,\Delta})}{\Delta}),$$

where we have used the smoothing property of expectation. However, the correlation in the user fingerprint vectors is weak, so that for any binary vector $s^n \in \{0,1\}^n$, we have

$$(1-\frac{|w_H(s^n)|}{\Delta})^{w_H(s^n)} \le \frac{\prod_{k=1}^n P_R(s_k)}{P_{(R_{i,j_k})_{k\in[n]}}(s^{n'})} \le (1+\frac{|w_H(s^n)|}{\Delta})^{w_H(s^n)},$$

where $w_H(\cdot)$ is the Hamming weight. Note that $w_H((R_{i,j_k})_{k\in[n]}) \to \mu$ with probability one due to concentration of measure. So, we conclude that $\frac{\prod_{k=1}^n P_R(s_k)}{P_{(R_{i,j_k})_{k\in[n]}}(s^{n'})} \approx 1$. The following proposition formalizes this statement. The proof is straightforward and is omitted for brevity.

Proposition 4 (Memoryless Fingerprints in SB): For a ground-truth graph generated according to the SB model, consider the partial fingerprint $\mathbf{R} \triangleq (R_{i,j_k})_{k\in[n']}, j_k \in [n], n' \in [n]$ of user $u_i, i \in [m]$. The following holds:

$$P_{\mathbf{R}}(s^{n'}) = o(1), s^{n'} \in \{0,1\}^n : w_H(s^{n'}) > \mu(1+\omega(1)),$$

Furthermore,

$$P_{\mathbf{R}}(s^{n'}) = (1+o(\frac{1}{n}))\prod_{k=1}^{n'} P_R(s_k),$$

as $n \to \infty$, where $s^{n'} \in \{0,1\}^n : w_H(s^{n'}) = \mu(1+O(1))$.

V. Sufficient Conditions for Successful Deanonymization

In this section, we derive sufficient conditions on the network parameters and the expected number of queries under which the attacker can successfully deanonymize the victim with vanishing probability of error as $m \to \infty$. Initially, we make simplifying assumptions on the scanning and querying noise statistics and develop the tools to study the more complex formulation in the next steps. We relax these assumptions in steps and derive general theoretical guarantees for successful deanonymization.

A. Identical Scanning Noise and Noiseless Query Responses

As a first step, we consider the scenario in which the scanning noise is identical for all users, i.e. $\Gamma = \{1\}$, and the query responses are received noiselessly, i.e. $|\Theta| = 1, P^1_{Y|E_0}(y|s) = \mathbb{1}(y = s), y, s \in \{0,1\}$.

Let us focus on the α-PA model for a given $\alpha \in (0,1]$. We generalize the information threshold strategy (ITS), which was introduced in [26], where we studied a scenario in which the ground truth is generated according to the IEE model. It was shown in [26] that the strategy is asymptotically optimal under IEE model — in terms of expected number of queries necessary for successful deanonymization with vanishing error. In the ITS, the attacker queries the group memberships of the victim starting from the first group r_1 and continuing by increasing the group index (i.e. $x_t = r_t, t \in [n]$), until a particular stopping criterion is met. To explain the stopping criterion, let us define the *information value* $I_k(t), k \in [m], t \in [n]$ of user u_k and time t as follows:

$$I_0(k) \triangleq \log P_M(k), \quad k \in [m], \tag{10}$$

$$I_t(k) \triangleq \sum_{i=1}^{t} \log \frac{P_{E_0|E_s}(y_i|f_{k,i})}{P_{E_0}(y_i)} + I_0(k), \quad k \in [m],\ t \in [n], \tag{11}$$

where $(f_{k,i})_{i\in[t]} \in \{0,1\}^t$ is the realization of the partial fingerprint of user u_k in the scanned graph

(i.e. $(F_{k,i})_{i\in[t]} = (f_{k,i})_{i\in[t]}$), the vector $y^t \in \{0,1\}^t$ is the realization of the vector of query responses (i.e. $Y^t = y^t$), and

$$P_{E_0|E_s}(y|f) \triangleq \frac{P_{E_0}(y)P^1_{E_s|E_0}(f|y)}{\sum_{y'\in\{0,1\}} P_{E_0}(y')P^1_{E_s|E_0}(f|y')}. \quad (12)$$

The concept of information values have also been used in evaluating hypothesis testing and communication with feedback scenarios in prior works (e.g. [37], [38]). Intuitively, the information value $I_t(k)$ captures the attacker's belief at time $t \in [n]$ about the possibility of user $u_k, k \in [m]$ being the victim, based on the received query responses, where a large positive $I_t(k)$ indicates a strong belief that the user is the victim, and a large negative $I_t(k)$ indicates a strong belief that the user is not the victim. For instance, the initial information value $I_0(k)$ captures the attacker's initial belief and is a large negative number if $P_M(k)$ is small. Furthermore, the terms $\log \frac{P_{E_0|E_s}(y_i|f_{k,i})}{P_{E_0}(y_i)}, i \in [t]$ in the definition of $I_t(k)$ in (10) capture the information gain of the attacker, after the query response at time t is received, about the possibility that user k is the victim. The identification function Id_t first determines whether the maximum information value of all users exceeds $\log\frac{1}{\epsilon}$, where the parameter $\epsilon > 0$ affects the resulting probability of error. If there exists a user whose information value exceeds $\log\frac{1}{\epsilon}$, that user is identified as the victim. Otherwise, the next query is made. So,

$$x_t(\mathcal{G}_s, Y^{t-1}) = r_t, \quad t \in [n] \quad (13)$$

$$Id_t(\mathcal{G}_s, Y^t) = \begin{cases} u_k & \text{if } \exists! k \in [m] : I_t(k) > log\frac{1}{\epsilon} \\ e & \text{Otherwise,} \end{cases} \quad t \in [n] \quad (14)$$

We call this attack strategy the ITS due to the use of information thresholds for deanonymization.

Theorem 1: Consider the ITS described above with parameter $\epsilon > 0$. Let $\overline{Q}_{ITS}$ be the resulting expected number of queries and $P_{e,ITS}$ the resulting probability of error. Then, in the α-PA scenario with $\alpha \in (0,1]$:

$$\overline{Q}_{ITS} \leq \frac{H(M) + \log\frac{1}{\epsilon} + i_{\max}}{c' I(E_0; E_s)}, \quad (15)$$

$$P_{e,ITS} \leq \frac{\epsilon}{c'}, \quad (16)$$

where c' is from Proposition 3, the mutual information is evaluated with respect to $P_{E_0,E_s} = P_{E_0}P_{E_s|E_0}$, the distribution $P_{E_s|E_0}$ is given in (12)*, the variable E_0 is Bernoulli with $P_{E_0}(1) = 1 - P_{E_0}(0) = \frac{\mu}{m}$, and $i_{max} \triangleq \max_{y,f\in\{0,1\}} \log \frac{P_{E_0|E_s}(y|f)}{P_{E_0}(y)}$.*

Proof: Appendix. □

At a high level, the bound on the expected number of queries necessary for successful deanonymization $\overline{Q}_{ITS}$ which is given in (15) can be interpreted as follows: the total initial uncertainty about the victim's index M is given by $H(M)$. The average information provided by each query response is $I(E_0; E_s)$, and the correlation in the group memberships induced in the generation process inflicts an information gain penalty captured by c'. That is, after $\overline{Q}_{ITS}$ queries, the attacker gains roughly $c'\overline{Q}_{ITS}I(E_0; E_s)$ bits of information regarding the victim's identity. Hence, successful identification occurs for $\overline{Q}_{ITS} \approx \frac{H(M)}{c'I(E_0;E_s)}$.

Remark 11: The coefficient c' in the denominator of $\frac{H(M)+\log\frac{1}{\epsilon}+i_{\max}}{c'I(E_0;E_s)}$ can be improved in special cases based on the value of α. For instance, it is shown in [26] that for the IEE model, where $\alpha \to 0$, the denominator $c'I(E_0; E_s)$ can be replaced by $I(E_0; E_s)$ to derive an asymptotically optimal bound.

Next, we focus on the SB model. Let $P^{\tau}_{E_0}(1) = \frac{\tau}{\sum_{\tau'\in\mathcal{T}} \tau'|\mathcal{C}_{\tau'}|}, \tau \in \mathcal{T}$, and let us assume without loss of generality that $P^1_{E_0}(1) \leq P^2_{E_0}(1) \leq \cdots \leq P^{|\mathcal{T}|}_{E_0} \leq \frac{1}{2}$. Then, the ITS query function queries the groups starting with most popular communities of groups. To elaborate, assume that $\mathcal{T} = \{1, 2, \cdots, |\mathcal{T}|\}$ and $\tau_0(j) \geq \tau_0(j'), j > j'$. Then $x(\mathcal{G}_s, Y^{t-1}) = r_t, t \in [n]$. Note that we have assumed that the attacker knows the community membership of the groups. In the absence of this information, the attacker may potentially extract the group's community memberships using $\mathcal{G}_s$.

The stopping criterion is modified as follows. The information value $I_k(t), k \in [m], t \in [n]$ of user u_k and time t is:

$$I_0(k) \triangleq \log P_M(k), \quad k \in [m], \quad (17)$$

$$I_t(k) \triangleq \sum_{\tau\leq\tau'} \sum_{\ell=1}^{|\mathcal{C}_\tau|} \log \frac{P_{E_0|E_s}(y_\ell|f_{k,\ell})}{P_{E_0^\tau}(y_\ell)} \quad (18)$$

$$+ \sum_{i=0}^{i'} \log \frac{P_{E_0|E_s}(y_i|f_{k,i})}{P_{E_0^{\tau'}}(y_i)} + I_0(k), \quad k \in [m],\ t \in [n] \quad (19)$$

where $t = \sum_{\tau\leq\tau'} |\mathcal{C}_\tau| + i', i' \leq |\mathcal{C}_{\tau'+1}|$. The initial information values $I_0(k), k \in [m]$ are defined in (17) in a similar fashion as that of (10). The definition of $I_t(k), k \in [m], t \in [n]$ in (11) is modified in (18) to capture the community-dependence of the edge probabilities in the SB model.

Theorem 2: In the SB scenario, let $\mathcal{T} = \{1, 2, \cdots, |\mathcal{T}|\}$ and assume that $\tau_0(j) \geq \tau_0(j'), j > j'$, then:

$$\overline{Q}_{ITS} \leq \sum_{\tau\leq\tau^*} |\mathcal{C}_\tau| + i^*$$

$$P_{e,ITS} \leq \epsilon,$$

where (τ^*, i^*) are defined as

$$\tau^* \triangleq \min_{\tau\in\mathcal{T}} \left\{\tau : \psi \leq \sum_{\tau'\leq\tau+1} |\mathcal{C}_\tau| I_\tau(E_0; E_s)\right\},$$

$$i^* \triangleq \min_{i\in[|\mathcal{C}_{\tau^*}|]} \left\{i : \psi \leq \sum_{\tau\leq\tau^*} |\mathcal{C}_\tau| I_\tau(E_0; E_s) + i I_{\tau^*+1}(E_0; E_s)\right\},$$

$$\psi \triangleq H(M) + \log\frac{1}{\epsilon} + i_{\max},$$

the mutual information $I_\tau(E_0;E_s)$ is evaluated with respect to $P^\tau_{E_0,E_s}=P^\tau_{E_0}P_{E_s|E_0}$, the variable E_0 is Bernoulli with parameter $P^\tau(E_0=1)=\frac{\tau}{\sum_{\tau'\in\mathcal{T}}\tau'|\mathcal{C}_{\tau'}|}\frac{\mu}{\beta}$, and $P_{E_s|E_0}$ is given in (12).

The proof for the upper bound on $\overline{Q}_{ITS}$ follows similar arguments as Theorem 1, and uses (16) along with the fact that

$$\mathbb{E}(I_{T_{n'}(M)})=\sum_{\tau\le\tau'}|\mathcal{C}_\tau|I_\tau(E_0;E_s)+i'I_{\tau'+1}(E_0;E_s),$$

where, $\mathbb{E}(T_{n'})=\sum_{\tau\le\tau'}|\mathcal{C}_\tau|+i',i'\le|\mathcal{C}|_{\tau'+1}$. That is, the average information gained after $T_{n'}$ queries is equal to $\mathbb{E}(I_{T_{n'}(M)})=\sum_{\tau\le\tau'}|\mathcal{C}_\tau|I_\tau(E_0;E_s)+i'I_{\tau'+1}(E_0;E_s)$. So, successful deanonymization requires an average of $\mathbb{E}(T_{n'})=\sum_{\tau\le\tau'}|\mathcal{C}_\tau|+i',i'\le|\mathcal{C}|_{\tau'+1}$ queries such that $\psi\le\sum_{\tau\le\tau'}|\mathcal{C}_\tau|I_\tau(E_0;E_s)+i'I_{\tau'+1}(E_0;E_s)$. The derivation of the upper bound on the error probability follows similar arguments as in the proof of Theorem 1 along with Proposition 4.

B. Noisy Scan and Query Responses

In this section, we extend the ITS to the case of noisy scan and query responses and arbitrary finite sets Γ and Θ. Note that as explained in Section III, the attacker does not have access to $\gamma(k),k\in[m]$ but has access to $\theta(M)$. To elaborate, the attacker knows the user device specifications, and hence it knows the query response noise statistics $P^{\theta(M)}_{Y|E_0}$, but does not know the users' privacy preferences, and hence it only knows that the scan knows statistics is given by one of the conditional distributions $P^\gamma_{E_s|E_0},\gamma\in\Gamma$, where γ may be different for different users.

Let us focus on the α-PA model for $\alpha\in(0,1]$. The query function is defined as in the previous section. The identification function is modified as follows. The attacker has access to $\theta(M)$ since it can query the victim's hardware and software specifications. So, it can find $P^{\theta(M)}_{Y|E_0}$ and use it in calculating the users' information values as in the previous scenario. As for the scan noise parameter, γ, the attacker computes $|\Gamma|$ different information values for each user, one for each value of $\gamma\in\Gamma$, and assigns the maximum resulting value as the information value of the user. This resembles the communication strategies used for communicating over compound channels when channel state information is unavailable [48]. So,

$$I_0(k)\triangleq\log P_M(k),\quad k\in[m],\tag{20}$$

$$I_t(k)\triangleq\max_{\gamma\in\Gamma}\sum_{i=1}^{t}\log\frac{P^{\gamma,\theta(M)}_{Y|E_s}(y_i|f_{k,i})}{P^\gamma_Y(y_i)}+I_0(k),\quad k\in[m],\quad t\in[n],\tag{21}$$

where $P^\gamma_Y(\cdot)=\sum_{s\in\{0,1\}}P_{E_0}(s)P^\gamma_{Y|E_0}(\cdot|s)$, and $P^{\gamma,\theta(M)}_{Y|E_s}(y_i|f_{k,i})=\sum_{s\in\{0,1\}}P^{\theta(M)}_{E_0|E_s}(s|f_{k,i})P^\gamma_{Y|E_0}(y|s),y,f_{k,i}\in\{0,1\}$, and:

$$P^{\Theta(M)}_{E_0|E_s}(s|f)\triangleq\frac{P_{E_0}(s)P^{\Theta(M)}_{E_s|E_0}(f|s)}{\sum_{s'\in\{0,1\}}P_{E_0}(s')P^{\Theta(M)}_{E_s|E_0}(f|s')}.\tag{22}$$

The initial information values in (20) are defined in a similar fashion as that of (10). The definition of $I_t(k),k\in[m],t\in[n]$ in (11) is modified in (21) to capture the effects of noise in query responses on the information gain at each step. The following sufficient conditions for successful deanonymization are given in the following theorem.

Theorem 3: Consider the ITS described above with parameter $\epsilon>0$. Let $\overline{Q}_{ITS}$ be the resulting expected number of queries and $P_{e,ITS}$ the resulting probability of error. Then, in the α-PA scenario with $\alpha\in(0,1]$:

$$\overline{Q}_{ITS}\le\sum_{\gamma\in\Gamma,\theta\in\Theta}P_{\Gamma,\Theta}(\gamma,\theta)\frac{H(M)+\log\frac{1}{\epsilon}+i_{\max}}{c'I_{\gamma,\theta}(Y;E_s)},\tag{23}$$

$$P_{e,ITS}\le\frac{|\Gamma|\epsilon}{c'},\tag{24}$$

where c' is from Proposition 3, the mutual information is evaluated with respect to $P_{E_0,E_s}=P_{E_0}P_{E_s|E_0}$, the distribution $P_{E_s|E_0}$ is given in (12)*, the variable E_0 is Bernoulli with $P_{E_0}(1)=1-P_{E_0}(0)=\frac{\mu}{m}$, $i_{max}\triangleq\max_{y,f\in\{0,1\}}\log\frac{P_{E_0|E_s}(y|f)}{P_{E_0}(y)}$, and $P_{\Gamma,\Theta}(\gamma,\theta)\triangleq\frac{|\{u_k|\theta(k)=\theta,\gamma(k)=\gamma\}|}{m},\theta\in\Theta,\gamma\in\Gamma$.*

The proof follows similar argument to that of Theorem 1. The derivation of the bound on $\overline{Q}_{ITS}$ for a given choice of $\theta\in\Theta,\gamma\in\Gamma$ is unchanged since the number of queries needed to achieve the desired information value threshold does not increase with the modified information values, since users are assigned a higher information value by maximizing over $\gamma\in\Gamma$. The bound on the probability of error follows the exact same steps as in the proof of Theorem 1 with the additional step of using the union bound to bound the probability of error over the union of choices of Γ. This leads to the addition of the coefficient $|\Gamma|$ in the upper bound on the probability of error in (24) as compared to the one given in (16).

Remark 12: Similar to the derivation of Theorem 3 which extends Theorem 1 to general scan and query noise statistics, Theorem 2 can also be extended to derive sufficient conditions for the success of ITS under the SB model and general noise statistics. Again, the bound on the expected number of queries $\overline{Q}$ remains the same, but the upper-bound on the probability of error P_e grows linearly in $|\Gamma|$.

VI. Simulation Results

In this section, we provide several simulations of synthesized and real-world ABND attacks to verify the theoretical results presented in the previous sections and gain further intuition regarding the users' privacy risks under such attack scenarios.

A. Effect of Growth Parameter α on $\overline{Q}_{ITS}$

As a first step, we consider a noiseless ABND scenario under the α-PA generation model, where the scanned graph and the query responses are acquired noiselessly by the attacker, and the victim is equally likely to be any of the users. We wish to evaluate the effect of changing the preferential attachment parameter $\alpha\in(0,1]$ on the expected number

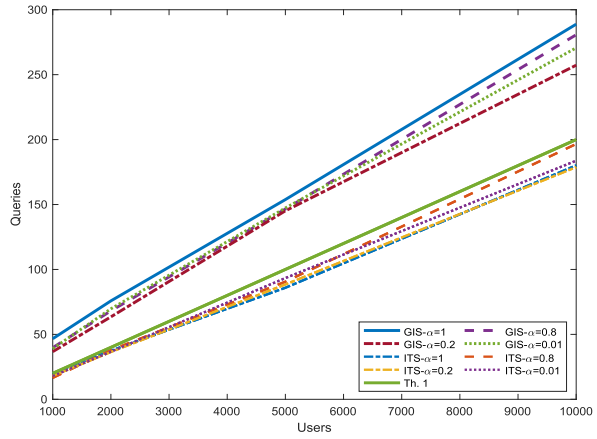


Fig. 3. Expected number of queries necessary for success in the ITS and the GIS under noiseless scanned and query response assumptions with a victim which is uniformly chosen among the users. The green (filled) line is the upper bound on the expected number of queries in ITS due to Theorem 1.

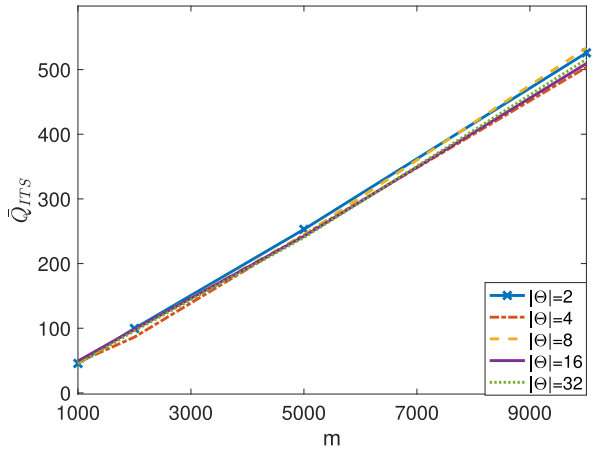


Fig. 4. Expected number of queries $\overline{Q}$ necessary for success in the ITS under noiseless scan and noisy query response assumptions with a victim which is uniformly chosen among the users.

of queries necessary for deanonymization under ITS. Note that in this case, ITS reduces to a simple strategy, where queries are made until the acquire responses have a unique match among the user fingerprints in the scanned graph since $P_{Y|E_s}(y|f) = \mathbb{1}(y = f), y, f \in \{0, 1\}$.

In order to provide a baseline for comparison, we also investigate the performance of a natural extension of the deanonymization strategy considered in [9] which is described below.

The Group Intersection Strategy (GIS): The attacker queries the group memberships of the victim sequentially starting with the membership in the first group, that is $x_t(\mathcal{G}_s, Y^{t-1}) = r_t, t \in [n]$. After receiving the query response at time $t \in [n]$, the attacker forms the ambiguity set $\mathcal{A}_t$ by intersecting the members of all groups $\mathcal{U}_i$ for which the query response is one, i.e. $\mathcal{A}_t = \bigcap_{i\in[t]:Y_i=1} \mathcal{U}_i, t \in [n]$. The attack concludes if $|\mathcal{A}_t| = 1$ in which case the unique user which is in $\mathcal{A}_t$ is declared as the victim.

In Figure 3, we have plotted the performance of ITS and GIS, where we have simulated the attack with parameters $\mu = 100$, $\epsilon = 0.01$, and $\beta = 0.1$. For each value $m = \{1000, 2000, 5000, 10000\}$, we have simulated the attack 500 times, by generating the ground-truth five times and choosing a victim randomly and uniformly for each generation 100 times. Our analysis in Theorem 1 predicts that $\overline{Q}_{ITS}$ grows linearly in $m \in \mathbb{N}$ since in the denominator in (15) we have $I(E_0; E_S) = H(E_0) = \frac{m}{\mu}(\log m + o(\log m))$, and in the numerator we have $H(M) = \log m$ and $i_{max} = \log m$. In fact, Theorem 1 predicts that $\overline{Q} \triangleq \overline{Q}_{ITS} - \frac{\log \frac{1}{\epsilon}}{\log m} \approx \frac{2m}{\mu}$, and does not depend on the value of α. This is verified by our simulations shown in Figure 3. Furthermore, it can be observed that the ITS outperforms the baseline GIS for all values of α nad m shown in Figure 3.

B. Effect of Query Response Noise on Q_{ITS}

Let us recall that the set Θ captures the diversity in query noise statistics due to the various hardware and software specifications of the users and the different browser sniffing techniques available to the attacker, where the resulting query response noise is captured by $P^{\theta(M)}_{Y|E_0}(\cdot|\cdot)$ and M is the victim's index. Now, we investigate the effect of diversity of query response noise on the expected number of queries for successful deanonymization with ITS. To elaborate, we consider a noiseless scanned graph but noisy query responses. To model the query noise diversity, we consider two initial noise statistics $P_{Y|E_0}$ and $P'_{Y|E_0}$, where $P_{Y|E_0}$ is the transition probability of a binary symmetric channel with parameter 0.01, and $P'_{Y|E_0}$ is the transition probability of a binary symmetric channel with parameter 0.3. These statistics are chosen to be within the range of empirical observations of noise in browser history sniffing (e.g. [13]–[15]). We consider 5 scenarios, where $\Theta_k = \{0, 1, 2, \cdots, 2^k - 1\}, k = 5$, and define $P^{\theta}_{Y|E_0} = \frac{\theta}{2^k-1} P_{Y|E_0} + \frac{2^k-1-\theta}{2^k-1} P'_{Y|E_0}, \theta \in \Theta_k$. Figure 4 shows the resulting expected number of queries as a function of m, where we have simulated the attack with parameters $\alpha = 1$, $\mu = 100$, $\epsilon = 0.1$, and $\beta = 0.4$. For each value $m = \{1000, 2000, 5000, 10000\}$, we have simulated the attack 500 times, by generating the ground-truth five times and choosing a victim randomly and uniformly for each generation 100 times. It can be observed that increasing users' query noise diversity does not have a significant effect on the probability of success.

C. Effect of Scanned Noise on $\overline{Q}_{ITS}$

In this section, we consider noiseless query responses, but noisy scanned graph and investigate the effects on the success of the ITS. As predicted by the theoretical results in Theorem 3, the diversity in scanned graph noise does not affect the expected number of queries. However, the upper-bound on the probability of error in Theorem 3 changes linearly in $|\Gamma|$. We have plotted the resulting probability of error in Figure 5, where in order to model the scan noise diversity, we have considered two initial noise statistics $P_{E_s|E_0}$ and $P'_{E_s|E_0}$, where $P_{E_s|E_0}$ is the transition probability of a binary

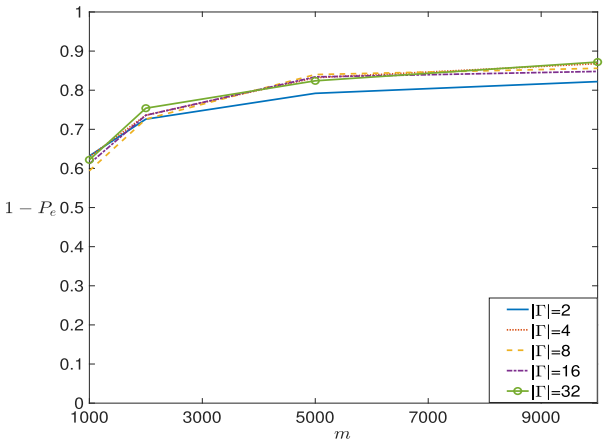
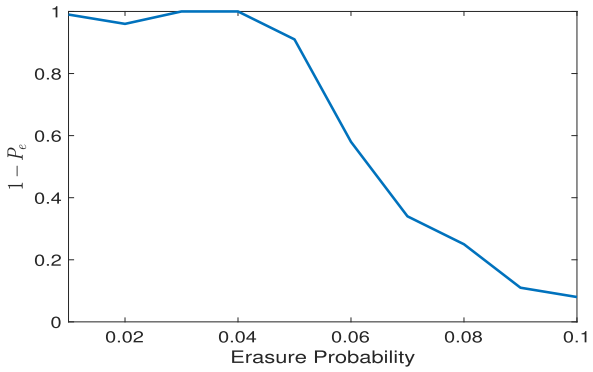
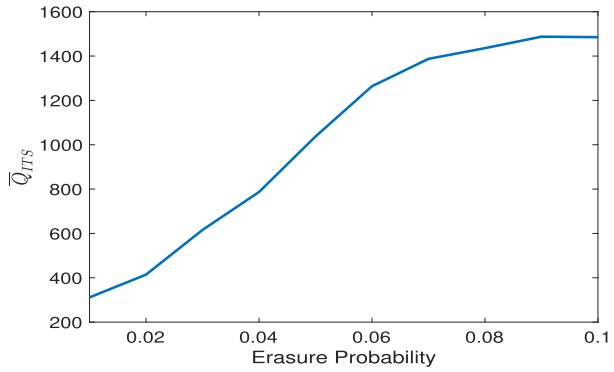


Fig. 5. Probability of success in the ITS under varying scanned noise statistics and noiseless query response assumptions with a victim which is uniformly chosen among the users.



(a)



(b)

Fig. 6. Probability of success (a) and average number of queries (b) for ITS in deanonymizing users in the LiveJournal network.

symmetric channel with parameter 0.01 and $P'_{E_s|E_0}$ is the transition probability of a binary symmetric channel with parameter 0.3. We have considered five scenarios, where $k \triangleq |\Gamma| = \{2, 4, 8, 16, 32\}$, and defined $P^{\gamma}_{E_s|E_0} = \frac{\gamma-1}{k-1} P_{E_s|E_0} + \frac{k-\gamma}{k-1} P_{Y|E_0}, \gamma \in [|\Gamma|]$. Figure 5 shows the resulting probability of success $(1 - P_e)$ as a function of m, where we have simulated the attack with parameters $\alpha = 1$, $\mu = 100$, $\epsilon = 0.1$, and $\beta = 0.4$. For each value $m = \{1000, 2000, 5000, 10000\}$, we have simulated the attack 500 times, by generating the ground-truth five times and choosing a victim randomly and uniformly for each generation 100 times. It can be observed that increasing the users' privacy preference options ($|\Gamma|$) does not have a significant effect on the resulting probability of success for ITS. This suggests that the upper-bound on the probability of error in Theorem 3 can be potentially improved.

D. Performance in Real-World Networks

In this section, we simulate an active attack on the LiveJournal network, which is a free on-line blogging community which allows users to form a group which other members can then join [49]. The database[2] consists of 3,997,962 members and 664,414 groups. We have extracted a subset of 1517 groups with at least 400 members, and selected a subset of 49,164 users which are members of at least 4 of these groups. The simulation is run 100 times, where each time a victim is chosen randomly and uniformly among the users. In Figure 6, we have simulated the attack under varying noise parameters, where we have modeled both the scanning and query noise with binary erasure channels with erasure probability ranging from 0.01 to 0.1. We have used $\epsilon = 0.1$ for the ITS error parameter. It can be seen that for the larger values of the erasure probability, the 1,517 groups scanned by the attacker are not sufficient to identify the victim and the attacker must scan and query additional group memberships, whereas for smaller erasure probability, the attacker succeeds with probability close to one.

VII. Conclusion and Future Work

We have studied the ABND problem for general non-equiprobable user indices under various ground truth generation models such as linear and sublinear preferential attachment and stochastic block model. We have studied the ITS deanonymization strategy which operates based on information thresholds. The strategy measures the amount of uncertainty in the user indices given the received query responses. We have characterized the performance of the ITS both for social networks with a fixed, finite number of users as well as for asymptotically large social networks. We have provided simulations of the attack both in synthesized as well as real-world bipartite networks to verify the theoretical results. Future research directions include i) extending the theoretical results to scenarios where the scan and query noise models allow for correlated noise, ii) exploring the model assumptions such as sparsity in real-world bipartite networks other than social networks such as wireless mobility, and medical databases, and iii) evaluating the performance of the proposed algorithms in such real-world bipartite networks.

[2]The database is available at https://snap.stanford.edu/data/com-LiveJournal.html.

APPENDIX
PROOF OF THEOREM 1

The proof builds upon ideas developed for studying the fundamental limit of communication over feedback channels [37]. Note that $I(E_0;E_s) \leq H(E_0) = (\frac{\mu}{m}\log m)(1+o(\frac{\log m}{m}))$ since $P(E_0=1)=\frac{\mu}{m}$. So, the upper-bound on $\overline{Q}$ is greater than $C_3\frac{n}{\mu}$ for some constant $C_3>0$. This allows us to use Proposition 3 to approximate the fingerprint distribution by a memoryless distribution as a lower bound. Define the following stopping times

$$\kappa_k \triangleq \min_{\kappa > \frac{C_3 n}{\mu}} \left\{\kappa \middle| I_k(\mathcal{G}_s, Y^\kappa) > \log\frac{1}{\epsilon}\right\}, \quad k\in[m],$$
$$\kappa^* \triangleq \min_{k\in[m]} \kappa_m$$

Note that by the definition of the identification function $Id_t(\cdot,\cdot)$ in (14), we have $\overline{Q}_{ITS} = \mathbb{E}(\kappa^*)$. We show that $\mathbb{E}(\kappa^*) \leq \frac{H(M)+\log\frac{1}{\epsilon}+i_{\max}}{I(E_0;E_s)}$. Note that $\mathbb{E}(\kappa^*) \leq \mathbb{E}(\kappa_M)$ by definition of κ^*. So, it is enough to prove the upper bound on $\mathbb{E}(\kappa_M)$. Fix $n'\in\mathbb{N}$. Let $T_{n'} = \min\{\kappa_M, n'\}$. Note that:

$$\begin{aligned}
\mathbb{E}\left(I_{T_{n'}}(M)\right) &= \mathbb{E}\left(\sum_{i=1}^{T_{n'}} \log\frac{P_{E_0|E_s}(Y_i|F_{k,i})}{P_{E_0}(Y_t)} + I_o(J)\right)\\
&\overset{(a)}{=} \mathbb{E}\left(\sum_{j=1}^{T_{n'}} \log\frac{P_{E_0|E_s}(Y_i|F_{k,i})}{P_{E_0}(Y_t)}\right) - H(J)\\
&\overset{(b)}{\geq} c' I(E_0;E_s)\mathbb{E}(T_{n'}) - H(J), \qquad (25)
\end{aligned}$$

where (a) uses linearity of expectation, and (b) follows from Wald's identity [50] and using $P((F_{k,i})_{i\in[t]}, Y^t) \leq c'\prod_{i=1}^t P_{E_0,E_s}(F_{k,i},Y_i), t\in\mathbb{N}$ which holds due to Proposition 3, and the fact that

$$\begin{aligned}
&P((F_{k,i})_{i\in[t]}, Y^t)\\
&= \sum_{s^t\in\{0,1\}^t} P((R_{k,i})_{i\in t} = s^t)\\
&\quad\times \prod_{i\in[t]} P_{E_s|E_0}(F_{k,i}|s_i)P_{Y|E_0}(Y_i|s_i)\\
&= \sum_{s^t\in\{0,1\}^t} P((R_{k,i})_{i\in t} = s^t)\prod_{i\in[t]} P_{E_s|E_0}(F_{k,i}|s_i)\mathbb{1}(Y_i=s_i)\\
&\leq c'\prod_{i\in[t]}\sum_{s_i\in\{0,1\}} P(R_{k,i}=s_i)P_{E_s|E_0}(F_{k,i}|s_i)\mathbb{1}(Y_i=s_i)\\
&= c'\prod_{i\in[t]} P_{E_0,E_s}(F_{k,i},Y_i).
\end{aligned}$$

Note that the sparsity condition $\sum_{i\in[t]} s_i = o(n)$ in Proposition 3 is satisfied with probability one due to Proposition 2. Also, note that at each step $t\in\mathbb{N}$, the increase in $I_t(M)$ is less than or equal to i_{max}. It follows that:

$$\mathbb{E}\left(I_{T_{n'}}(M)\right) \leq \mathbb{E}\left(I_{T_{n'}-1}(M)\right) + i_{max} \leq \log\frac{1}{\epsilon} + i_{max}, \qquad (26)$$

where we have used the fact that and that by the definition of κ_M and $T_{n'}$, we have $I_{T_{n'}-1}(M) \leq \log\frac{1}{\epsilon}$ since $T_{n'}-1 < \kappa_M$. Combining (25) and (26) we get $\mathbb{E}(T_{n'}) \leq \frac{H(M)+\log\frac{1}{\epsilon}+i_{\max}}{c' I(E_0;E_s)}$, and using the monotone convergence theorem by increasing n' asymptotically, we get $\mathbb{E}(T_{n'}) = \mathbb{E}(\kappa_M) = \overline{Q}_{ITS}$ which yields the desired bound on $\overline{Q}_{ITS}$. It remains to prove the bound on $P_{e,ITS}$. We have:

$$\begin{aligned}
P_e &= P(\exists j\neq M : \kappa_j \leq \kappa_M) \leq \sum_{j\neq M} P(\kappa_j \leq \infty)\\
&= \sum_{j\neq M}\lim_{\eta\to\infty} P(\kappa_j\leq\eta)\\
&\overset{(a)}{=} \sum_{j\neq M}\lim_{\eta\to\infty} \mathbb{E}_{P_{Y^n,(F_{M,i})_{i\in[n]}}}\\
&\quad\times\left(\frac{P_{Y^n}P_{(F_{M,i})_{i\in[n]}}}{P_{Y^n,(F_{M,i})_{i\in[n]}}}\mathbb{1}(\kappa_j\leq\eta))\right)\\
&\leq \sum_{j\neq M}\lim_{\eta\to\infty}\frac{(1+o(1))}{c'}\mathbb{E}_{P_{Y_i,F_{M,i}}}\\
&\quad\times\left(\prod_{i\in[n]}\frac{P_{Y_i}P_{F_{M,i}}}{P_{Y_i,F_{M,i}}}\mathbb{1}(\kappa_j\leq\eta))\right)\\
&= \sum_{j\neq M}\lim_{\eta\to\infty}\frac{(1+o(1))}{c'}\mathbb{E}_{P_{Y_i,F_{M,i}}}\\
&\quad\times\left(\prod_{i\in[\eta]}\frac{P_{Y_i}P_{F_{M,i}}}{P_{Y_i,F_{M,i}}}\mathbb{1}(\kappa_j\leq\eta))\right)\\
&\quad\times E_{P_{Y_i,F_{M,i}}}\left(\prod_{i\in[\eta+1,n]}\frac{P_{Y_i}P_{F_{M,i}}}{P_{Y_i,F_{M,i}}}\right)\\
&\quad\times\sum_{j\neq M}\lim_{\eta\to\infty}\frac{(1+o(1))}{c'}\mathbb{E}_{P_{Y_i,F_{M,i}}}\\
&\quad\times\left(\prod_{i\in[\eta]}\frac{P_{Y_i}P_{F_{M,i}}}{P_{Y_i,F_{M,i}}}\mathbb{1}(\kappa_j\leq\eta))\right)\\
&= \sum_{j\neq M}\lim_{\eta\to\infty}\frac{(1+o(1))}{c'}\mathbb{E}_{P_{Y_i,F_{M,i}}}\\
&\quad\times\left(e^{\sum_{i\in[\eta]}\log\frac{P_{Y_i}P_{F_{M,i}}}{P_{Y_i,F_{M,i}}}}\mathbb{1}(\kappa_j\leq\eta))\right)\\
&\leq \sum_{j\neq M}\lim_{\eta\to\infty}\frac{(1+o(1))}{c'}\mathbb{E}_{P_{Y_i,F_{M,i}}}\\
&\quad\times\left(e^{\sum_{i\in[\eta]}\log\frac{P_{Y_i}P_{F_{M,i}}}{P_{Y_i,F_{M,i}}}}\right)\\
&= \sum_{j\neq M}\lim_{\eta\to\infty}\frac{(1+o(1))}{c'}\mathbb{E}_{P_{Y_i,F_{M,i}}}\left(e^{-I_\eta(M)-I_0(M)}\right)\\
&\leq \sum_{j\neq M}\lim_{\eta\to\infty}\frac{(1+o(1))}{c'}\mathbb{E}_{P_{Y_i,F_{M,i}}}\left(e^{-\log\frac{1}{\epsilon}-I_0(M)}\right)\\
&= \sum_{j\neq M}\frac{1}{c'}\epsilon P_M(j) \leq \frac{1}{c'}\epsilon(1+o(1)).
\end{aligned}$$

where in (a) we have used the fact that $P_{(F_{j,i})_{i\in[n]}} = P_{(F_{M,i})_{i\in[n]}}, j\in[m]$.

REFERENCES

[1] J. Estrada-Jiménez, J. Parra-Arnau, A. Rodríguez-Hoyos, and J. Forné, "Online advertising: Analysis of privacy threats and protection approaches," *Comput. Commun.*, vol. 100, pp. 32–51, Mar. 2017.

[2] A. Razaghpanah *et al.*, "Apps, trackers, privacy, and regulators: A global study of the mobile tracking ecosystem," in *Proc. Netw. Distrib. Syst. Secur. Symp.*, 2018, pp. 1–8.

[3] P. Mavriki and M. Karyda, "Using personalization technologies for political purposes: Privacy implications," in *Proc. Int. Conf. E-Democracy*. Cham, Switzerland: Springer, 2017, pp. 33–46.

[4] N. Takbiri, A. Houmansadr, D. L. Goeckel, and H. Pishro-Nik, "Limits of location privacy under anonymization and obfuscation," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 764–768.

[5] Z. Montazeri, A. Houmansadr, and H. Pishro-Nik, "Achieving perfect location privacy in wireless devices using anonymization," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 11, pp. 2683–2698, Nov. 2017.

[6] N. Takbiri, R. Soltani, D. L. Goeckel, A. Houmansadr, and H. Pishro-Nik, "Asymptotic loss in privacy due to dependency in Gaussian traces," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2019, pp. 1–6.

[7] Y.-A. de Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel, "Unique in the crowd: The privacy bounds of human mobility," *Sci. Rep.*, vol. 3, no. 1, Dec. 2013, Art. no. 1376.

[8] V. D. Blondel, A. Decuyper, and G. Krings, "A survey of results on mobile phone datasets analysis," *EPJ Data Sci.*, vol. 4, no. 1, p. 10, Dec. 2015.

[9] G. Wondracek, T. Holz, E. Kirda, and C. Kruegel, "A practical attack to de-anonymize social network users," in *Proc. S&P*, May 2010, pp. 223–238.

[10] M. Fire, R. Goldschmidt, and Y. Elovici, "Online social networks: Threats and solutions," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 2019–2036, 4th Quart., 2014.

[11] J. Su, A. Shukla, S. Goel, and A. Narayanan, "De-anonymizing web browsing data with social networks," in *Proc. 26th Int. Conf. World Wide Web*, Apr. 2017, pp. 1261–1269.

[12] J. Domingo-Ferrer, D. Sánchez, and J. Soria-Comas, "Database anonymization: Privacy models, data utility, and microaggregation-based inter-model connections," *Synth. Lectures Inf. Secur., Privacy, Trust*, vol. 8, no. 1, pp. 1–136, Jan. 2016.

[13] B. Gulmezoglu, A. Zankl, T. Eisenbarth, and B. Sunar, "Perfweb: How to violate web privacy with hardware performance events," in *Proc. Eur. Symp. Res. Comput. Secur.* Cham, Switzerland: Springer, 2017, pp. 80–97.

[14] M. Smith, C. Disselkoen, S. Narayan, F. Brown, and D. Stefan, "Browser history re: Visited," in *Proc. 12th Workshop Offensive Technol.*, 2018, pp. 1–13.

[15] A. Shusterman *et al.*, "Robust website fingerprinting through the cache occupancy channel," in *Proc. 28th Secur. Symp.*, 2019, pp. 639–656.

[16] D. Irani, S. Webb, K. Li, and C. Pu, "Large online social footprints—An emerging threat," in *Proc. Int. Conf. Comput. Sci. Eng.*, 2009, pp. 271–276.

[17] M. Jakobsson and S. Stamm, "Invasive browser sniffing and countermeasures," in *Proc. 15th Int. Conf. World Wide Web*, 2006, pp. 523–532.

[18] M. Balduzzi *et al.*, "Abusing social networks for automated user profiling," in *Proc. RAID*, 2010, pp. 422–441.

[19] L. Bilge, T. Strufe, D. Balzarotti, and E. Kirda, "All your contacts are belong to us: Automated identity theft attacks on social networks," in *Proc. 18th Int. Conf. World Wide Web*, 2009, pp. 551–560.

[20] M. Srivatsa and M. Hicks, "Deanonymizing mobility traces: Using social network as a side-channel," in *Proc. ACM Conf. Comput. Commun. Secur.*, 2012, pp. 628–637.

[21] S. Kinsella, V. Murdock, and N. O'Hare, "'I'm eating a sandwich in Glasgow' modeling locations with tweetss," in *Proc. 3rd Int. Workshop Search Mining User-Generated Contents*, New York, NY, USA, 2011, pp. 61–68.

[22] Z. Cheng, J. Caverlee, and K. Lee, "You are where you tweet: A content-based approach to geo-locating Twitter users," in *Proc. 19th ACM Int. Conf. Inf. Knowl. Manage.*, 2010, pp. 759–768.

[23] D. Gruss, D. Bidner, and S. Mangard, "Practical memory deduplication attacks in sandboxed javascript," in *Proc. Eur. Symp. Res. Comput. Secur.* Cham, Switzerland: Springer, 2015, pp. 108–122.

[24] H. Kim, S. Lee, and J. Kim, "Inferring browser activity and status through remote monitoring of storage usage," in *Proc. 32nd Annu. Conf. Comput. Secur. Appl.*, Dec. 2016, pp. 410–421.

[25] F. Shirani, S. Garg, and E. Erkip, "An information theoretic framework for active de-anonymization in social networks based on group memberships," in *Proc. 55th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2017, pp. 470–477.

[26] F. Shirani, S. Garg, and E. Erkip, "Optimal active social network de-anonymization using information thresholds," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 1445–1449.

[27] A. Capocci *et al.*, "Preferential attachment in the growth of social networks: The internet encyclopedia Wikipedia," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 74, no. 3, 2006, Art. no. 036116.

[28] M. E. J. Newman, "Clustering and preferential attachment in growing networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 64, no. 2, 2001, Art. no. 025102.

[29] R. Albert, "Scale-free networks in cell biology," *J. Cell Sci.*, vol. 118, no. 21, pp. 4947–4957, 2005.

[30] V. Borrel, M. D. De Amorim, and S. Fdida, "A preferential attachment gathering mobility model," *IEEE Commun. Lett.*, vol. 9, no. 10, pp. 900–902, Oct. 2005.

[31] R. Van Der Hofstad. (2009). *Random Graphs and Complex Networks*. [Online]. Available: http://www.win.tue.nl/rhofstad/NotesRGCN.pdf

[32] M. Bloznelis and M. Karoński, "Random intersection graph process," *Internet Math.*, vol. 11, nos. 4–5, pp. 385–402, 2015.

[33] H. A. Simon, *Models of Man: Social and Rational*. Hoboken, NJ, USA: Wiley, 1957.

[34] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, Oct. 1999.

[35] L. Florescu and W. Perkins, "Spectral thresholds in the bipartite stochastic block model," in *Proc. Conf. Learn. Theory*, 2016, pp. 943–959.

[36] T. C. Yen and D. B. Larremore, "Community detection in bipartite networks with stochastic block models," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 102, no. 3, 2020, Art. no. 032309.

[37] M. V. Burnashev, "Data transmission over a discrete channel with feedback. Random Transmission time," *Problemy Inf. Inform.*, vol. 12, no. 4, pp. 10–30, Dec. 1976.

[38] M. Naghshvar and T. Javidi, "Active sequential hypothesis testing," *Ann. Statist.*, vol. 41, no. 6, pp. 2703–2738, 2013.

[39] J. Kunegis, M. Blattner, and C. Moser, "Preferential attachment in online networks: Measurement and explanations," in *Proc. 5th Annu. ACM Web Sci. Conf.*, 2013, pp. 205–214.

[40] F. Peruani, M. Choudhury, A. Mukherjee, and N. Ganguly, "Emergence of a non-scaling degree distribution in bipartite networks: A numerical and analytical study," *Europhys. Lett. (EPL)*, vol. 79, no. 2, p. 28001, Jul. 2007.

[41] M. Peltomäki and M. Alava, "Correlations in bipartite collaboration networks," *J. Stat. Mech., Theory Exp.*, vol. 2006, no. 1, Jan. 2006, Art. no. P01010.

[42] L. Olejnik, C. Castelluccia, and A. Janc, "Why Johnny can't browse in peace: On the uniqueness of web browsing history patterns," in *Proc. 5th Workshop Hot Topics Privacy Enhancing Technol.*, 2012, pp. 1–17.

[43] P. Erdős and A. Rényi, "On random graphs I," *Mathematicae*, vol. 6, pp. 290–297, Oct. 1959.

[44] E. Zheleva, H. Sharara, and L. Getoor, "Co-evolution of social and affiliation networks," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2009, pp. 1007–1016.

[45] M. M. Christiansen, K. R. Duffy, F. du Pin Calmon, and M. Médard, "Multi-user guesswork and brute force security," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6876–6886, Dec. 2015.

[46] M. Shariatnasab, F. Shirani, and E. Erkip, "Fundamental privacy limits in bipartite networks under active attacks," 2021, *arXiv:2106.04766*.

[47] R. Impagliazzo and V. Kabanets, "Constructive proofs of concentration bounds," in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*. Berlin, Germany: Springer, 2010, pp. 617–631.

[48] H. Boche, R. F. Schaefer, and H. Vincent Poor, "On the ε-capacity of finite compound channels with applications to the strong converse and second order coding rate," in *Proc. 54th Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2020, pp. 1–6.

[49] J. Yang and J. Leskovec, "Defining and evaluating network communities based on ground-truth," *Knowl. Inf. Syst.*, vol. 42, no. 1, pp. 181–213, Jan. 2015, doi: 10.1007/s10115-013-0693-z.

[50] A. Wald, "On cumulative sums of random variables," *Ann. Math. Statist.*, vol. 15, no. 3, pp. 283–296, Sep. 1944.

Mahshad Shariatnasab (Graduate Student Member, IEEE) received the B.Sc. degree from Shahid Beheshti University in 2015 and the M.Sc. degree from the Khajeh Nasir Toosi University of Technology in 2017. She is currently pursuing the Ph.D. degree with the Electrical and Computer Engineering Department, North Dakota State University. Her research interests are in the general areas of security and privacy. Her recent work includes developing information theoretic methods for analysis of fundamental limits of web privacy.

Farhad Shirani (Member, IEEE) received the B.S. degree in electrical engineering from the Sharif University of Technology, and the M.Sc. degree in applied mathematics and the M.Sc. and Ph.D. degrees from the University of Michigan, Ann Arbor, MI, USA. He served as a Lecturer and a Post-Doctoral Research Fellow at the University of Michigan in 2017. He was a Research Assistant Professor at New York University from 2017 to 2020. He is an Assistant Professor at the Electrical and Computer Engineering Department, North Dakota State University. His research interests include privacy and security, wireless communications, and machine learning. His recent work includes developing information theoretic methods for analysis of fundamental limits of web privacy, design of receiver architectures for energy efficient communication over MIMO systems, and design of algorithms for opportunistic multi-user scheduling under various fairness constraints.

Elza Erkip (Fellow, IEEE) received the B.S. degree in electrical and electronics engineering from Middle East Technical University, Ankara, Turkey, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, USA.

She is an Institute Professor at the Electrical and Computer Engineering Department, New York University Tandon School of Engineering. Her research interests are in information theory, communication theory, and wireless communications.

Dr. Erkip is a member of the Science Academy of Turkey and is a Clarivate Highly Cited Researcher. She has been a member of the Board of Governors of the IEEE Information Theory Society since 2012, where she was the Society President in 2018. She was a Distinguished Lecturer of the IEEE Information Theory Society from 2013 to 2014. She received the NSF CAREER Award in 2001, the IEEE Communications Society WICE Outstanding Achievement Award in 2016, and the IEEE Communications Society Communication Theory Technical Committee (CTTC) Technical Achievement Award in 2018. Her paper awards include the IEEE Communications Society Stephen O. Rice Paper Prize in 2004 and the IEEE Communications Society Award for Advances in Communication in 2013. She has had many editorial and conference organization responsibilities. Some recent ones include the General Co-Chair of the IEEE International Symposium of Information Theory in 2013, the Track Chair of the Asilomar Conference on Signals, Systems and Computers, and the MIMO Communications and Signal Processing in 2017, the Technical Co-Chair of the IEEE Wireless Communications and Networking Conference in 2017, and the Guest Editor of IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS in 2015.