Path-preserving Anonymization for Inter-domain Routing Policies

Xiaozhe Shao, Hossein Pishro-Nik, and Lixin Gao

Department of Electrical and Computer Engineering, University of Massachusetts Amherst
{xiaozheshao, pishro, lgao}@engin.umass.edu

Abstract. The Internet consists of tens of thousands of autonomous systems (ASes), which exchange routing information through policy-based routing protocols, such as Border Gateway Protocol (BGP). BGP allows network administrators/operators to configure routing policies independently. The routing result is derived from complex interactions controlled by the configured policies. Thus, the routing policies of all networks are essential to understand which network a packet uses to reach a destination or whether there is a path to reach a destination. However, the routing policies of a network are proprietary information. The majority of networks in the Internet do not publicly share their routing policies. In this paper, we anonymize the routing policy with the aim to preserve the utility to understand the routing system. To this end, we perturb both the routing policies and the AS-level topology. We propose a series of anonymization schemes that perturb the AS-level topology to derive a k-anonymity graph. We further perturb routing policies to preserve the best path. We evaluate the anonymization schemes against sophisticated de-anonymization attacks exploiting rich structural information of AS-level topologies, such as neighborhood and reference distance. The experimental result shows that the de-anonymization attack exploiting rich structural information can identify less than 1% of nodes.

Keywords: Graph Anonymization · K-anonymity · Network Verification

1 Introduction

The Internet consists of tens of thousands of autonomous systems (ASes), each of which belongs to an organization such as an Internet Service Provider (ISP), a university, a company, or an Internet Exchange Point (IXP). The interconnections in the Internet have a hierarchical structure, where tier-1 ISPs provide settlement-free services among each other, and provide transit services to regional ISPs or stub networks, and regional ISPs provide transit services to even smaller ISPs or stub networks. BGP is a policy-based protocol used for routing among ASes.

The route taken by a packet to traverse in the Internet is determined by the route selection process, and a result of complex interactions controlled by the configured routing policies of all networks. Knowing routing policies of all networks allow us to answer a broad set of what-if questions [18, 11]. For example, if a network wants the inbound traffic from Google to go through a specific provider, then before changing its routing configuration the operator wants to verify that the configuration indeed leads to the desired outcome.

Routing policies are considered to be proprietary by their networks. Each network administrator/operator configures its routing policy independently. The routing policy of a network reveals how the network exchanges routes with neighbors. Thus, the commercial agreement (*e.g.*, provider-customer relationships) between neighboring ASes can be inferred through their routing policies. Further, the routing policy of a network reveals that which neighbor is more preferred than others. Therefore, routing policies should be anonymized before sharing.

A trivial method of anonymizing routing policies is to anonymize IDs of all networks, *e.g.* AS Numbers, in the routing policies. Namely, each original AS Number is transformed into a unique ID. However, anonymizing IDs alone is not sufficient. The number of neighbors of a network can be inferred from the anonymized routing policy of the network, even if all network IDs contained in the routing policy are anonymized.

In the Internet, tier-1 ISPs have a large number of neighbors (customers and peers) and the number of neighbors follow power-law distribution [12]. The ASes with a large number of neighbors are few and their degrees (*e.g.* the number of neighbors) are usually unique. The AS-level topology can be derived from the publicly available routing tables of RouteViews [14] and RIPE RIS [9]. The degrees in the AS-level topology can be used to identify the AS. Even if the derived AS-level topology does not exactly reflect the AS-level topology in reality, the ranking of these large ISPs in terms of degree will be same. Thus, these Tier-1 ISPs can still be identified. Identifying a few tier-1 networks can lead to identifying tier-2 networks by what tier-1 neighbors they have. Progressively, more and more ASes might be identified when their higher-tier neighbors have been identified. Therefore, anonymizing the AS-level topology is the key in anonymizing routing policies.

Anonymizing graphs have been studied at the area of online social networks [6, 20, 21, 2, 10, 8, 3, 17, 19]. These schemes perturb the graph through adding and/or removing nodes and/or edges in social networks, so that the adversaries can not identify users from the social network. These schemes focus on preserving social network structural properties. However, these schemes do not maintain the best paths derived from the routing policies. The applications exploiting the inter-domain routing policy are very sensitive to the best paths derived from the routing policy. Therefore, these schemes can not be directly applied to the anonymization of routing policies.

In this paper, we propose a path-preserving anonymization scheme for sharing inter-domain routing policies. We anonymize the routing policies so that the best paths can be maintained after anonymization. To do that, we first perturb AS-level topology by adding fake networks and connection between fake networks and between fake and real networks. We then establish routing policies for these newly-added links.

To perturb the AS-level topology, one anonymization scheme is adding neighborhood relationships between pairs of networks. However, simply adding fake edges between networks is not sufficient to maintain the best path utility. The reason is that the local preferences on the fake edge between two networks can not be properly assigned. When network u is connected to network v through a fake edge, to avoid the fake edge being included in the best paths, all real edges should be preferred over the fake edge. Thus, network u and network v should give each other the lowest local preference. In practice, a network assigns the lowest local preference to its provider and a pair of neighboring ASes can not be the provider of each other. Therefore, the fake edges between a pair of ASes, such as network u and network v, can be identified by using the local preferences assigned to the fake edge.

We propose to add fake networks into the AS-level topology and then add the connection between fake networks and real networks to perturb the AS-level topology. In the topology perturbation, we first consider to avoid being identified through the number of neighbors. We propose a deterministic anonymization scheme which derives a k-anonymity graph. As a result, the probability of identifying a network by the degree information is at most $\frac{1}{k}$. However, the resulting k-anonymity graph is deterministic. We further propose the probabilistic anonymization scheme to generate the k-anonymity graph with higher randomness. To anonymize the AS-level topology, the probabilistic anonymization scheme might significantly enlarge the graph with fake networks. Adding too many fake networks might make the scheme infeasible in practice. Finally, we propose the hybrid anonymization scheme to take advantage of the strengths of both the deterministic anonymization scheme and the probabilistic anonymization scheme.

After fake networks and fake edges are added, we propose to generate the associated routing policies for those fake networks and edges. To avoid de-anonymization, in the anonymized routing policies, the local preferences assigned to real neighboring ASes and fake neighboring ASes have the same distribution.

We evaluate the proposed anonymization schemes against de-anonymization attacks exploiting rich structural information, such as neighborhood and reference distance. The experimental result shows that the structure-based de-anonymization attacks can identify less than 1% networks from the anonymized AS-level topology. In addition, we use Closeness Centrality as an example metrics to illustrate how hard to identify a network through the structural information. We also show that the forged routing policies of networks added by the anonymization schemes will not be used to identify the added networks from the original networks.

The rest of this paper is organized as follow: In Section 2, we illustrate background and formally define the problem. In Section 3, we discuss the preliminaries necessary for preserving utility to construct the anonymization schemes. To perturb the AS-level topology, in Section 4 we propose the deterministic anonymization scheme, the probabilistic anonymization scheme and the hybrid anonymization scheme respectively. In Section 5, we propose to perturb the routing policies through a local preference assignment scheme. We describe the experiment in Section 6, related work in Section 7, and make the conclusion of the paper in Section 8.

2 Background and Problem Formulation

2.1 Routing Policies and AS-level Topology

BGP is the de-facto standard for inter-domain routing in the Internet. BGP allows network operators to configure the routing policy of their own networks. The routing policies of those networks together determine the best routes from source to destination networks. In BGP, a network assigns a local preference to each of its neighbors respectively and prefers the routes from a neighbor network with a larger local preference.

Listing 1.1: A snippet of the routing policy

```
router bgp 45000
   neighbor 192.168.1.2 remote-as 40000
   neighbor 192.168.1.2 route-map SET-LOCAL-PREF-CUST in
   neighbor 192.168.1.3 remote-as 50000
   neighbor 192.168.1.3 route-map SET-LOCAL-PREF-PEER in
!
route-map SET-LOCAL-PREF-CUST
   set local-preference 200
route-map SET-LOCAL-PREF-PEER
   set local-preference 100
```

In Listing 1.1, we illustrate a snippet of the routing policy for AS 45000. The routing policy of AS 45000 indicates the neighborhood relationships of AS 45000. According to the routing policy, AS 45000 has two neighboring ASes, AS 40000 and AS 50000. This snippet also indicates that AS 45000 sets up the local preference 200 for its customer, AS 40000, and the local preference 100 for its peer, AS 50000. As a result, AS 45000 prefers routes from AS 40000 over that from AS 50000.

The routing policy of a network indicates the neighboring ASes of the network. The routing policies of all networks together implies the AS-level topology of the Internet. Let us denote the AS-level topology by a graph, $G = \{V, E, R\}$, where V is the set of nodes, E is the set of edges and R is the set of all routing policies. Each node represents an AS in the Internet. The edge between two nodes represents the link between two networks.

2.2 Path-preserving Anonymization

To protect the routing policies, we can obfuscate the key information in the routing policies. One of the most trivial one is pseudonymization. We can replace AS numbers in the routing policies by a pseudo ID, so that the routing policies can not be identified through the AS numbers directly. Let us denote the pseudonymization as a network mapping function, π, that maps the original AS number to the anonymized AS number (the pseudo ID).

Anonymizing IDs only is not enough for the anonymization of routing policies. The neighboring relationship with the real preference and the real neighbor ID implied by the routing policies can be used for identification as we state in the introduction. Beyond pseudonymization, we might also change the neighboring ASes and local preferences for these neighboring ASes in a routing policy. As for the routing policy in Listing 1.1, we might switch the local preferences of AS 40000 and AS 50000. However, the obfuscation might not preserve the best paths derived by the routing policies. For example, after switching the local preferences in the routing policy of Listing 1.1, AS 45000 prefers routes from AS 50000 over that from AS 40000.

Answering what-if questions by exploiting the inter-domain routing policy relies on the best paths derived from the routing policies. If the best paths of the anonymized routing policy is different from the best paths of the original routing policy, then the anonymized routing policy might not be useful to answer what-if questions. In order to preserve the utility of the routing policy, a best path derived by the original routing policy should be a best path derived by the anonymized routing policy. We say that the anonymization scheme preserves the *best path utility*, if, for an AS-level topology $G = \{V, E, R\}$, $\forall i, j \in V$, $\pi(Path(i, j)) = Path(\pi(i), \pi(j))$, where $Path(i, j)$ indicates the best path from network i to network j and $\pi(Path(i, j))$ maps all original AS numbers of ASes in the best path to their anonymized AS numbers.

In this paper, we propose the anonymization schemes that preserve the best path utility. To do so, all networks and the neighboring relationships are maintained after the routing policy anonymization. Namely, the anonymization schemes should not remove nodes or edges from the AS-level topology. Instead, the anonymization schemes can add new nodes and new edges into the AS-level topology. In addition, the local preferences of neighboring ASes can be changed but the ranking of these local preferences should be maintained. Thus, after the anonymization, a network ranks the neighboring ASes in the same way as before the anonymization.

2.3 Threat Model

After the anonymization, the routing policies of all networks are publicly available. The adversary can access the anonymized routing policies of all networks and try to figure out the routing policy for each network. Given an AS number, the adversary needs to identify the routing policy of this AS and discovers how this AS ranks its neighboring ASes from the routing policy. To preserve the best path utility, how an AS ranks its neighboring ASes is maintained in the anonymized routing policy. Therefore, the key of this de-anonymization procedure is to infer the network mapping function, π.

The AS-level topology implied by the anonymized routing policies can be used to identify the routing policy of each network. The routing tables of Route-Views [14] and RIPE RIS [9] are publicly available. The adversary can get the AS-level topology of the Internet by using the routing tables. To preserve the best path utility, all networks and their connections are maintained in the anonymized AS-level topology. We consider that the adversary tries to use the AS-level topology of the Internet to identify the networks in the anonymized AS-level topology and figure out the routing policy of the networks.

3 Anonymization Scheme Overview

To anonymize the routing policies, we first consider how to perturb the AS-level topology implied by the routing policies, so that the networks are hard to be identified through the number of neighbors. Then, we modify the routing policies of these networks accordingly, so that the networks are hard to be identified through the local preferences in the anonymized routing policies.

3.1 Anonymizing Graph Structure

To perturb the AS-level topology, we consider to generate *k-anonymity* graphs. A k-anonymity graph is a graph that, for any node, there are at least $k-1$ other nodes sharing the same degree with the node. If only the degree information is used by the adversary, the probability that the adversary can correctly identify a real node from the k-anonymity graph is at most $\frac{1}{k}$.

An anonymized graph that is not k-anonymity can also guarantee that the adversary can not correctly identify a real node with the probability higher than $\frac{1}{k}$. Following this idea, we extend the definition of k-anonymity graph as follows. Given an anonymized graph derived from the original graph, the anonymized graph is a k-anonymity graph, if the probability that the adversary can correctly identify a real node is no more than $\frac{1}{k}$ when the original graph size goes to infinity. Formally, $\forall u \in V$ and $\forall v \in V'$,

$$\lim_{|V|\to\infty} P\big(P(\pi(u) = v|\mathbf{Y} = \mathbf{y}) \leq \frac{1}{k}\big) = 1, \tag{1}$$

where V and V' are the sets of nodes in the original graph and the anonymized graph respectively and $\mathbf{y}$ is the degree sequence of nodes in the anonymized graph. In Section 4, we propose three graph anonymization schemes to generate k-anonymity graph.

Although k-anonymity with a small k is vulnerable to de-anonymization attacks, the schemes proposed in this paper can select a large k, such as k equals the number of real nodes. The reason is that the schemes in this paper preserve the best path utility. Selecting a large k does not impact the utility of the anonymized graph. In the experiment, we will show that selecting a large k can achieve pretty good anonymization performance.

To preserve the best path utility, all nodes and edges in the original graph should be maintained in the anonymized graph. Therefore, the anonymization schemes should not remove nodes or edges. In this paper, we propose to add fake nodes into the graph. Then, fake nodes and real nodes are connected by a set of fake edges to perturb the graph structure.

3.2 Local Preference Assignment for Fake Edges

After the graph anonymization, we need to modify all routing policies for newly-added nodes and edges. For each fake node added into the topology, we need to create a new routing policy for the node. For each fake edge added into the topology, the local preferences should be properly assigned to the edge. In Section 5, we propose a local preference assignment scheme to determine local preferences for these fake edges, so that the fake nodes can not be identified through their local preferences.

4 Node Anonymization

In this section, we propose three anonymization schemes to perturb the AS-level topology. In Section 4.1, we first propose the deterministic anonymization scheme which generates a k-anonymity graph. However, the resulting graph is determined by the Internet topology and lacks of randomness. To increase the randomness of the resulting graph, we propose the probabilistic anonymization scheme in Section 4.2. The probabilistic anonymization scheme might significantly enlarge the graph with fake networks. Finally, in Section 4.3, we propose the hybrid anonymization scheme that takes the advantages of both the deterministic and probabilistic anonymization schemes.

4.1 Deterministic Anonymization Scheme

In the deterministic anonymization scheme, we partition all nodes of the original graph into several groups, where each group contains at least k nodes. Then, we add fake edges between these real nodes and fake nodes to make these real nodes in the same group have the same degree. Finally, fake edges are added between fake nodes to make all fake nodes have the same degree.

4.1.1 Node Grouping Given a graph with n nodes, we partition all real nodes into $\lfloor \frac{n}{k} \rfloor$ groups, where each group at least has k nodes. To do so, we sort the nodes of the original graph in the descending order of their degrees. For the first $\lfloor \frac{n}{k} \rfloor - 1$ group, the group i is composed of the $(i * k - k + 1)$-th node to the $(i * k)$-th node. The last group is composed of the $(\lfloor \frac{n}{k} \rfloor * k - k + 1)$-th node to the n-th node.

For each group, we select a target degree. All real nodes in the group are supposed to reach the target degree of this group after the anonymization. To increase the degree of real nodes, we add fake nodes and fake edges. We also select a target degree for all fake nodes. When every node reaches their target degrees, we get a k-anonymity graph. Please refer to the technical report [1] for how we select the target degrees and the number of fake nodes added into the anonymized graph.

4.1.2 Graph Construction Given the number of fake nodes and the target degrees, we add fake edges to generate k-anonymity graph. We define *degree gap* of a node as the target degree of this node minus its degree. We use g_i to denote the degree gap of node i. The degree of fake nodes in the original graph are zero. Through adding fake edges, we will make the degree gaps of all nodes be zero. To do so, we select one node with the largest degree gap and make the node reach its target degree in each step. We show the details as follows.

Increasing real node degree To increase the degree of real nodes, we always select the real node that has the largest degree gap. Suppose that we select node i with the degree gap g_i. Then, we select g_i fake nodes that have the largest degree gap and connect these g_i fake nodes with node i with fake nodes.

Increasing fake node degree To increase the degree of fake nodes, we always select the fake node that has the largest degree gap. Suppose that we select node i with the degree gap g_i. Then, besides node i, we select g_i fake nodes that have the largest degree gap and connect these g_i fake nodes with node i with fake nodes.

4.1.3 Privacy Guarantee

Theorem 1. *The deterministic anonymization scheme can generate a k-anonymity graph.*

Please refer to the technical report for the proof of this Theorem.

4.2 Probabilistic Anonymization Scheme

The deterministic scheme lacks of uncertainty. The regularity of the anonymized graph generated by the deterministic scheme can be used for de-anonymization. In order to increase the uncertainty of the anonymization scheme, we propose probabilistic anonymization scheme. In probabilistic scheme, we add randomness during the graph anonymization so that the resulting graph is not deterministic.

4.2.1 Graph Anonymization The basic idea of the graph anonymization is to connect a real node and a fake node with the probability, p. For each real node, the number of its fake neighbors is a random variable that follows binomial distribution with the average as $m * p$ and the variance as $m * p(1 - p)$.

[1] Please access to the technical report through http://rio.ecs.umass.edu/mnilpub/papers/crisis2021-shao-tr.pdf

When the difference between the degrees of two real nodes in the original graph is insignificant comparing to $m * p(1-p)$, these two real nodes can not be identified through their degrees.

Given a graph G, we refer to a group with k nodes in G as *k-node group.* In a graph G, we define *k-node degree delta* of a k-node group as the difference between the maximal and the minimal degrees of nodes in the group. Then, for a specific node, there are a number of k-node groups that contain this node. We define *minimal k-node degree delta* of a node as the minimal k-node degree delta of k-node groups that contain the node. We use $\delta_{(i,k)}$ to denote the minimal k-node degree delta of node i. We refer to that k-node group as minimal k-node group of this node. We construct the k-anonymity graph as follows.

Adding Fake Nodes We add $m = max(\delta^2_{(max,k)} \ln n, n)$ fake nodes into the graph, where $\delta_{(max,k)}$ is the maximal $\delta_{(i,k)}$ for $\forall i \in [1, n]$, n is the number of nodes in G and d_{max} is the maximal degree in the graph G. We will show that m is sufficient in Theorem 2.

Connecting Real Nodes with Fake Nodes For each real node, we add a fake edge between this real node and each fake node with a probability p, where $0 < p < 1$. On average, each real node will connect to $m * p$ fake nodes.

Connecting Fake Nodes with Fake Nodes Then, we connect fake nodes with fake nodes. The goal is to make the expected degree of fake nodes be the expected degree of the real node with the highest degree in the original graph. To do that, each fake node connects to another fake node with the probability $q = \frac{(m-n)p+d_{max}}{m-1}$.

4.2.2 Probabilistic Privacy Guarantee

Theorem 2. *The probabilistic scheme generates a k-anonymity graph.*

We provide the proof of this Theorem in the technical report. In the following, we just illustrate the intuitive idea that the anonymization method can generate a k-anonymity graph. In the anonymization procedure, a number of fake edges are added to each real and fake node. With a higher probability (the probability goes to one when the graph size goes to infinity), the k nodes in the minimal k-node group of each node will have very similar degrees in the anonymized graph. Because, the minimal k-node degree delta will be asymptotically smaller than the variance for the number of the fake edges connected to each real node. Apparently, it is more likely that the node with higher degree in original graph has the higher degree in the anonymized graph. However, when the degree of a node in the anonymized graph is dominated by the fake edges added in the anonymization procedure, the degree delta of the nodes in the original graph are insignificant. Then, the k nodes in each group have similar degrees. The probability of successfully identifying nodes by their degrees is $1/k$.

4.3 Hybrid Anonymization Scheme

In the probabilistic anonymization scheme, the number of fake nodes needed to perturb the graph rises rapidly with the increasing of the degree difference. In order to reduce the number of added faked nodes, we propose the hybrid anonymization scheme. We will first use the deterministic anonymization scheme to reduce the degree difference in the graph. After that we derive probabilistic k-anonymity graph by using the probabilistic anonymization scheme.

4.3.1 Graph Construction Hybrid anonymization scheme guarantees to generate a k-anonymity graph. The key idea is to reduce the degree difference through the technique of the deterministic anonymization scheme and then apply the probabilistic anonymization scheme. We describe the two phases as follows.

Reducing Degree Delta through Deterministic Anonymization Scheme Given a graph with n nodes, to reduce the degree delta, we first partition the nodes into $\lfloor \frac{n}{k} \rfloor$ groups, where each group at least has k nodes. In this phase, our aim is to make sure that the degree delta of each group is less than a target degree delta, δ_{target}. We select $\delta_{target} = \lfloor \sqrt{n/\ln n} \rfloor$. When $\delta_{max,k}$ is less than δ_{target}, in a graph with n nodes, according to Theorem 2, n fake nodes are enough to generate a k-anonymity graph in the next phase. The detailed steps of this phase is similar to the deterministic anonymization scheme.

Perturbing Node Degrees through Probabilistic Anonymization Scheme In this phase, we randomly increase the node degrees by using the same technique of the probabilistic anonymization method. We treat both the real nodes and the fake nodes added by the last phase as the real nodes. Then, we can use the probabilistic anonymization scheme to anonymize the graph.

4.3.2 Algorithm Analysis

Theorem 3. *The hybrid anonymization scheme generates a k-anonymity graph.*

Proof. When the degree delta of each group is reduced into $\lfloor \sqrt{n/\ln n} \rfloor$, the minimal k-node degree delta of each node is equal to or less than $\lfloor \sqrt{n/\ln n} \rfloor$. According to Theorem 2, we can prove this theorem.

5 Routing Policy Anonymization

In this section, we propose a local preference assignment scheme for fake edges. We aim to guarantee that the real nodes and fake nodes have the similar local preferences, so that fake nodes can not be identified by the local preferences. We consider both *outgoing preference* and *incoming preference*. The outgoing preference of a node is the average of local preferences that are assigned to its neighbors by this node. The incoming preference of a node is the average of local preferences that are assigned to this node by its neighbors. The proposed scheme generates local preferences so that the distribution of the outgoing preferences and the incoming preferences for real nodes are same with that for fake nodes.

The best paths are determined by the ranking of the local preferences instead of the exact value of the local preferences. In this paper, we use the ranking number of these local preferences to represent the original values of these local preferences. For example, if a node assigns local preferences to its four neighbors as 100, 200, 200 and 400, then the associated ranking numbers are 3, 2, 2 and 1 respectively.

5.1 Accommodating the Distribution of Outgoing Preference

To adjust the outgoing preferences of a node, we need to determine ranking number for neighbors. Namely, how many neighbors have the ranking number 1, how many neighbors have the ranking number 2 and so on. We refer to the set of neighbors with the ranking number i as the *i-th ranking set* of the node. The union of all ranking sets of a node should be the neighbor set of the node. Note, in this step, we determine how many fake edges will be assigned with a ranking number instead of assigning the ranking number to specific fake edges. In the following, we illustrate how we determine the size of the ranking sets for each node.

In each step, we randomly select a node and determine the size of the ranking sets. For each node, we initiate the procedure from determining the size of first ranking set and finish the procedure until the total size of ranking sets equals the number of neighbors. To preserve the best path utility, a real node will not change the local preference for its real neighbors and will not prefer fake neighbors over real neighbors. Namely, the top ranking sets of a real node are already determined by the original routing policy. For these real nodes, we select the size of ranking sets for their fake neighboring relationships only.

To determine the size of i-th ranking set for a node, we investigate the size of i-th ranking sets of the other nodes. If the size of i-th ranking sets of the other nodes are all zero, then we just pick the size for i-th ranking set of this node. Therefore, the total size of ranking sets of this node equals to the node degree. Otherwise, we select an existing size for i-th ranking set of this node. More specifically, the probability of selecting the size of i-th ranking set as s is as follows.

$$p(s,i) = \begin{cases} 0 & \forall k, N(k,i) = 0 \\ \frac{1/N(s,i)}{\sum_{(k|N(k,i)\neq 0)} 1/N(k,i)} & N(s,i) \neq 0. \end{cases} \tag{2}$$

where s is the size of a ranking set and $N(s,i)$ is the number of i-th ranking sets with the size s.

5.2 Accommodating the Distribution of Incoming Preference

Given the size of all ranking sets, we select specific neighbors for each ranking set. The goal is to make the incoming preference of real nodes have the same distribution with that of fake nodes. To preserve the best path utility, real nodes always prefer their real neighbors over fake neighbors. To avoid fake nodes having large incoming preference, we make fake nodes prefer fake neighbors over real neighbors.

The maximal degree of real nodes, d_{max}, can be used by the adversary to remove fake edges. For a node i, only if node j is one of its top d_{max} neighbors in the ascending order of local preferences, then the edge (i, j) could be a real edge. Otherwise, the edge (i, j) has to be a fake edge. To avoid an edge (i, j) from being identified as a fake edge, node i and node j should treat each other as top d_{max} neighbors. To do so, before assigning the local preference value to each fake edge, for each node, we select a set of neighbors as its top d_{max} neighbors. We refer to those neighbors of the node as *high-ranking neighbors.* Apparently, for a node, there are at most d_{max} high-ranking neighbors. The rest neighbors are *low-ranking neighbors.* Only the edges between the node and its low-ranking neighbors are identified as fake edges.

6 Experiment

6.1 Datasets

To evaluate the proposed anonymization schemes, we anonymize a series of graphs through these schemes. Since Internet topologies are power-law graphs, we synthesize power-law graphs with various sizes for evaluation. In order to generate power-law graphs that have the same properties with Internet topologies, we exploit the graph generation algorithm in [4]. We also use the Internet AS-level topologies in CAIDA AS Relationship Database [1] to evaluate the performance of these schemes on Internet topologies.

6.2 Robustness against De-anonymization Algorithm

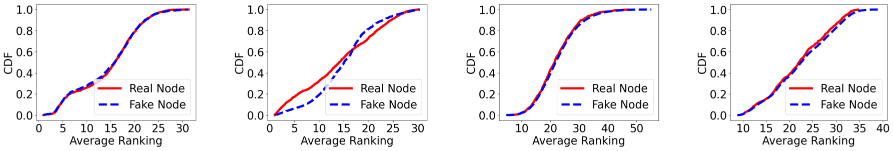
There are a number of de-anonymiztion algorithms which exploit structural information to identify nodes from the anonymized graph. Most of them are seed-based. Those algorithms rely on a seed graph which already maps a subset of real nodes in the original graph to the nodes in the anonymized graph. However, in the context of de-anonymizing network graph, it is hard to get a set of seed nodes. Therefore, in the experiment, we consider the seedless de-anonymization algorithm to identify the real nodes in the anonymized graph. We exploit the state-of-the-art seedless de-anonymization algorithm [5] to identify the nodes from the anonymized graphs.

Table 1 illustrates the number of real nodes identified through de-anonymization algorithm. In this experiment, we synthesize power-law graphs of various size and perform the anonymization and de-anonymization algorithms. For each size, we generate five graphs and illustrate the average number of real nodes that are correctly identified in Table 1. Although the anonymization schemes aim to perturb the degree of each nodes, to some extent, they can defend the de-anonymization algorithm using structural information. When k is n, $\sqrt{n}$ or $\ln n$, more than 90% real nodes can not be correctly identified.

Table 1: Percentage of real nodes that are identified by seedless de-anonymization algorithm [5].

Number of Nodes	100	400	700	$1,000$
Hybrid n-anonymity	0%	0%	0%	0%
Hybrid $\sqrt{n}$-anonymity	0.2%	0.1%	0.03%	0%
Hybrid $\ln n$-anonymity	1.2%	0.15%	0%	0.1%
Deterministic n-anonymity	0%	0%	0%	0.02%
Deterministic $\sqrt{n}$-anonymity	2.4%	1.05%	0.94%	0.92%
Deterministic $\ln n$-anonymity	7.2%	5.1%	7.31%	7.28%

6.3 Potential De-anonymization Using Local Preference



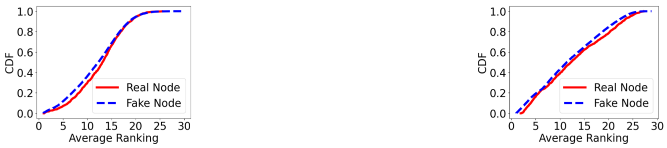
(a) Outgoing Preference (DAS). (b) Incoming Preference (DAS). (c) Outgoing Preference (HAS). (d) Incoming Preference (HAS).

Fig. 1: Distribution of local preferences with Deterministic Anonymization Scheme (DAS) and Hybrid Anonymization Scheme (HAS).

The local preference of nodes in the anonymized graph might be used to separate real and fake nodes. In this experiment, we investigate the distribution of outgoing and incoming preferences of real nodes and fake nodes. When the distribution for real nodes are similar with that for fake nodes, the local preference is hard to be used to identify real nodes.

We illustrate CDF of outgoing and incoming after deterministic and hybrid anonymization schemes in Figure 1 and Figure 2 respectively. As Figure 1 shows, the distributions for real and fake nodes are very similar. The adversary might or might not use the local preference and the maximal degree to identify fake edges. Namely, for each AS, only the d_{max} top-ranking links can be real links while the other links have to be fake. These fake edges can be removed. Thus, in the experiments, we consider the scenarios where the adversary uses or does not use fake edge removal. As Figure 2 shows, the distributions for real and fake nodes are still very similar after removing fake edges.

Fake edges in the anonymized graph can be removed through observing the outgoing local preference of each node. Even after removing fake edges, the seedless de-anonymization attack can not identify real nodes from the anonymized graphs. The result is similar to Table 1. Due to the space limit, we do not list the result.



(a) Outgoing Preference.

(b) Incoming Preference.

Fig. 2: Distribution of local preferences with Hybrid Anonymization Scheme after fake node removal.

7 Related Work

7.1 Internet-Wide Network Verification

A number of formal methods have been exploited to verify properties for interdomain routing. To study the safety property of BGP systems, Satisfiable Module Theories (SMT) [16] and Rewriting Logic [15] are used to verify the convergence conditions. A model checking tool is used to search possible attraction attacks on the Internet [13]. A policy-aware model [11] is proposed for routing verification at Internet scale. Those verification schemes need both the Internet topologies and the routing policies of networks for the verification. To preserve the confidentiality of routing policies, the anonymization schemes proposed in this paper can generate anonymized graphs for those verification schemes.

7.2 Graph Anonymization

A series of graph anonymization mechanisms have been proposed at the area of online social networks [6, 20, 21, 2, 10, 8, 3, 17, 19]. One large class of anonymization mechanisms exploits k-anonymity, where k nodes in the anonymized graph can not be separated from each other. The first approach in this direction is proposed in [6], where each node shares the same degrees with at least $k-1$ other nodes. At the same time, Zhou and Pei [20] propose the concept of k-neighborhood anonymity to defend against neighborhood attacks. To defend against more sophisticated structural attacks, k-automorphism [21] and k-isomorphism [2] are proposed. Our schemes follow k-anonymity as well as preserve the best path utility for routing verification.

Another class of anonymization mechanisms use the Differential Privacy technique to provide strong privacy guarantee [10, 8, 3, 17, 19]. Those mechanisms generate a synthetic graph which maintains structural similarity to the original graph. However, the best paths of the synthetic graph will be different from that of the original graph, since there is not an injective function mapping from nodes of the original graph to nodes of the synthetic graph.

To preserve edge privacy, Mittal et al. proposed a Random Walk based anonymization technique [7]. Although the preserving edge privacy can avoid local preferences from being revealed, the best paths in the anonymized graph are changed and the graph utility can not be preserved.

8 Conclusion

In this paper, we anonymize the inter-domain routing policies. We design the anonymization schemes to guarantee the policy privacy of networks without jeopardizing the best path utility of the routing policies. To perturb the Internet topology, we propose deterministic, probabilistic and hybrid anonymization schemes. These three schemes guarantee that degree-based de-anonymization can not be identified network from the anonymized graph with a probability higher than $\frac{1}{k}$. Even if the de-anonymization attack using structural information can not identify more than 1% of networks. After topology anonymization, we anonymize the routing policies accordingly. To do so, we propose a local preference assignment scheme. The resulting local preferences in the anonymized graph preserve the best path utility and can not be used to identify networks.

9 Acknowledgement

This work was supported in part by NSF grants CNS-1900866 and CCF-1918187, and the Radcliffe fellowship. We thank Catherine Greenhill for many helpful discussions.

References

1. CAIDA: The caida as relationships dataset, <1998-2020> (Oct 2019), http://www.caida.org/data/as-relationships/
2. Cheng, J., Fu, A.W.c., Liu, J.: K-isomorphism: privacy preserving network publication against structural attacks. In: Proceedings of the 2010 ACM SIGMOD International Conference on Management of data. pp. 459–470 (2010)
3. Dwork, C., McSherry, F., Nissim, K., Smith, A.: Calibrating noise to sensitivity in private data analysis. In: Theory of cryptography conference. pp. 265–284. Springer (2006)
4. Elmokashfi, A., Kvalbein, A., Dovrolis, C.: On the scalability of bgp: The role of topology growth. IEEE Journal on Selected Areas in Communications **28**(8), 1250–1261 (2010)
5. Ji, S., Li, W., Srivatsa, M., Beyah, R.: Structural data de-anonymization: Quantification, practice, and implications. In: Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security. pp. 1040–1053. CCS '14, Association for Computing Machinery, New York, NY, USA (2014). https://doi.org/10.1145/2660267.2660278, https://doi.org/10.1145/2660267.2660278
6. Liu, K., Terzi, E.: Towards identity anonymization on graphs. In: Proceedings of the 2008 ACM SIGMOD international conference on Management of data. pp. 93–106 (2008)
7. Mittal, P., Papamanthou, C., Song, D.: Preserving link privacy in social network based systems. arXiv preprint arXiv:1208.6189 (2012)
8. Proserpio, D., Goldberg, S., McSherry, F.: A workflow for differentially-private graph synthesis. In: Proceedings of the 2012 ACM workshop on Workshop on online social networks. pp. 13–18 (2012)

9. RIPE NCC: Routing information service. https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris, https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris
10. Sala, A., Zhao, X., Wilson, C., Zheng, H., Zhao, B.Y.: Sharing graphs using differentially private graph models. In: Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference. pp. 81–98 (2011)
11. Shao, X., Gao, L.: Verifying policy-based routing at internet scale. In: IEEE INFOCOM 2020 - IEEE Conference on Computer Communications. pp. 2293–2302 (2020)
12. Siganos, G., Faloutsos, M., Faloutsos, P., Faloutsos, C.: Power laws and the as-level internet topology. IEEE/ACM Trans. Netw. **11**(4), 514âĂŞ524 (Aug 2003). https://doi.org/10.1109/TNET.2003.815300, https://doi.org/10.1109/TNET.2003.815300
13. Sosnovich, A., Grumberg, O., Nakibly, G.: Analyzing internet routing security using model checking. In: Proceedings of the 20th International Conference on Logic for Programming, Artificial Intelligence, and Reasoning - Volume 9450. pp. 112–129. LPAR-20 2015, Springer-Verlag New York, Inc., New York, NY, USA (2015). https://doi.org/10.1007/978-3-662-48899-7_9, https://doi.org/10.1007/978-3-662-48899-7_9
14. University of Oregon: Route Views. www.routeviews.org (2021), www.routeviews.org
15. Wang, A., Gurney, A.J.T., Han, X., Cao, J., Loo, B.T., Talcott, C., Scedrov, A.: A reduction-based approach towards scaling up formal analysis of internet configurations. In: IEEE INFOCOM 2014 - IEEE Conference on Computer Communications. pp. 637–645 (April 2014). https://doi.org/10.1109/INFOCOM.2014.6847989
16. Wang, A., Jia, L., Zhou, W., Ren, Y., Loo, B.T., Rexford, J., Nigam, V., Scedrov, A., Talcott, C.: Fsr: Formal analysis and implementation toolkit for safe interdomain routing. IEEE/ACM Trans. Netw. **20**(6), 1814–1827 (Dec 2012). https://doi.org/10.1109/TNET.2012.2187924, https://doi.org/10.1109/TNET.2012.2187924
17. Wang, Y., Wu, X.: Preserving differential privacy in degree-correlation based graph generation. Transactions on data privacy **6**(2), 127 (2013)
18. Weitz, K., Woos, D., Torlak, E., Ernst, M.D., Krishnamurthy, A., Tatlock, Z.: Scalable verification of border gateway protocol configurations with an smt solver. In: Proceedings of the 2016 ACM SIGPLAN International Conference on Object-Oriented Programming, Systems, Languages, and Applications. pp. 765–780. OOPSLA 2016, Association for Computing Machinery, New York, NY, USA (2016). https://doi.org/10.1145/2983990.2984012, https://doi.org/10.1145/2983990.2984012
19. Xiao, Q., Chen, R., Tan, K.L.: Differentially private network data release via structural inference. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 911–920 (2014)
20. Zhou, B., Pei, J.: Preserving privacy in social networks against neighborhood attacks. In: 2008 IEEE 24th International Conference on Data Engineering. pp. 506–515. IEEE (2008)
21. Zou, L., Chen, L., Özsu, M.T.: K-automorphism: A general framework for privacy preserving network publication. Proceedings of the VLDB Endowment **2**(1), 946–957 (2009)