More Anonymity Through Trust Degree in Trust-based Onion Routing

Peng Zhou, Xiapu Luo, and Rocky K. C. Chang

Department of Computing, The Hong Kong Polytechnic University, Hunghom, Hong Kong
{cspzhouroc,csxluo,csrchang}@comp.polyu.edu.hk

Abstract. Trust-based onion routing employs users' own trust to circumvent compromised onion routers. However, it runs a high risk of being deanonymized by the inference attacks based on a priori trust relationships. In this paper, we first observe that the onion routers with higher trust degrees (e.g., those that are trusted by more other users) are more effective in defending against the inference attacks. We therefore incorporate trust degree into trust-based onion routing. With a rigorous theoretical analysis, we devise an optimal strategy for router selection and an optimal routing algorithm for path selection. Both minimize the risk of deanonymization through inference without sacrificing the capability of evading compromised routers. Moreover, simulation-based experiments on top of real-world social networks confirm the effectiveness of the optimal router selection.

Key words: trust degree, anonymity, trust-based onion routing

1 Introduction

Recently, trust-based models have attracted growing research interests in the anonymous communication area [1–4], especially for onion routing [5–7]. Onion routing networks protect anonymity with the help of onion routers. However, since onion routers are usually deployed by volunteers whose identities and technical competence are not verified [7], users loose the chance to detect the compromised routers. And even worse, various attacks employ compromised routers to deanonymize users [8–20]. The most recent research proposes trust-based onion routing algorithms to address this problem [2, 4]. By considering the trust that an user assigns to routers' owner, this user can select routers from trusted individuals, hence circumventing the compromised routers.

In existing trust-based onion routing networks, the user only considers its own trust and believes that routers with equal trust can protect its anonymity equivalently. However, if an adversary can observe the routers in a user's connection and make inference based on the knowledge of a priori trust relationships, the user is more likely to be deanonymized if she selects the routers that are rarely trusted by other users. As studied in [4, 21], this inference attack is a major threat to trust-based onion routing. Therefore, besides the user's own trust for router selection, the trust from other users also plays a very important role in protecting anonymity. We find that the routers trusted by more other users are more effective in defending against the inference attack. In this paper, we define a router's trust degree with respect to a user as the sum of trust from other users in this router.

Figure 1 illustrates the effectiveness of routers' trust degree in protecting anonymity. In this example, users can only select their trusted onion routers to make their connections. Alice equivalently trusts Bob and Ken, both of whom operate onion routers. Pete is an adversary who knows the trust relationships among the users and routers. If Pete observes Bob's router in Alice's connection, he can deanonymize Alice directly as Bob is only trusted by Alice. However, Pete cannot deanonymize Alice by observing Ken's router, because Ken is also trusted by many other users.

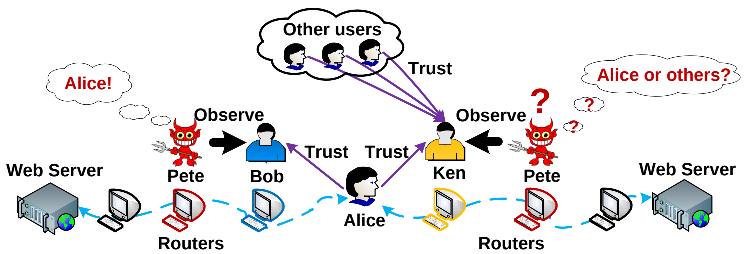


Fig. 1. Trust degree in protecting anonymity.

Moreover, since we observe in the real world that each person's friends are always trusted by different number of other people, an average person can potentially gain more anonymity by considering trust degree in trust-based onion routing networks. To conclude this observation, we analyze a public data set from the Facebook reported in [22]. This data set regards other people in a person's friend list as friends with equal trust. The authors of this data set crawled the New Orleans regional network in Facebook from December 29th, 2008 to January 3rd, 2009 and collected more than 1.5 millions social links from about 60 thousands persons to their friends. Out of them, $53,609$ persons have more than one friend.

Figure 2 illustrates the distribution of trust degrees of these $53,609$ persons' friends in [22]. We calculate a friend's trust degree with respect to a person as the number of other persons who have this friend in their friend lists. The horizontal axis represents the person index while the vertical axis shows the trust degree of persons' friends. For the ease of explanation, we sort these persons in an ascending order according to their trust degree distance, which can be computed by subtracting the smallest trust degree from the largest one of each person's friends. It can be seen, more than 99.6% persons have friends with different trust degree. In particular, for more than 80% persons, their friends' trust degree varies larger than 50.

Trust degree is an intuitive, but effective, feature in defending against a-priori trust relationship based inference, but is neglected in prior research. By selecting routers with a large trust degree, the user can intelligently hide its identity with the help of many other users, hence obtain more protection for its anonymity. In this paper, we incorporate trust degree into the trust-based onion routing. In particular, we make three major contributions:

1. To the best of our knowledge, we are the first to incorporate the trust from other users into the trust-based onion routing.

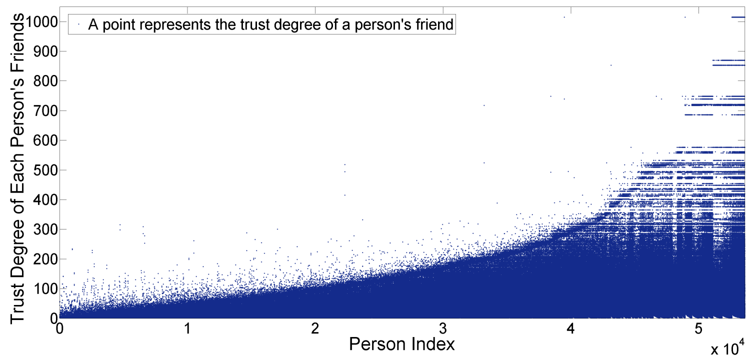


Fig. 2. Trust degree distribution of persons' friends.

2. More importantly, we prove an optimal router selection strategy based on the trust from other users. This minimizes the chance of deanonymization through inference, but does not sacrifice the capability of evading compromised routers. We evaluate this strategy in both simulation and real-world social networks. Experimental results show the user anonymity can be effectively improved.
3. We also prove an optimal trust degree aware routing algorithm for path selection.

The remainder of this paper is organized as follows. We review related works in Section 2. Section 3 introduces necessary preliminaries, including the trust model, the adversary model, the definition of trust degree, and the anonymity. In Section 4, we present an optimal strategy for router selection that incorporates trust degree. We also analyze the anonymity improvement in both simulation and real-world social networks. In Section 5, we prove an optimal trust degree aware routing algorithm for path selection in theory. We conclude this paper in Section 6.

2 Related Work

Trust-based onion routing appears recently and attracts growing interests from both industrial and academic communities [1–4]. Trust is effective in identifying compromised routers [2,4], and thus defending against correlation-like attacks [8–20]. However, users who select routers according to trust run a high risk to be deanonymized by the adversary who knows a priori trust relationships [4, 21].

In this section, we review three kinds of past work in the literature. We first present a brief description of the attacks that rely on compromised routers. We then review existing trust-based anonymous communications. Moreover, we also discuss the side effect if the trust models are used to protect anonymity.

In onion routing networks, users anonymously access the Internet through layered encrypted circuits. These circuits are established by dynamically selected onion routers [5–7]. However, without an effective mechanism to verify routers' identities, onion routing networks are vulnerable to compromised routers. A number of attacks exploit compromised routers to compromise anonymity in onion routing networks. This includes the predecessor attack [8], the congestion attack [9], the traffic analysis attacks [10, 11, 23, 24], the sybil attack [12] and many other correlation attacks [13–20].

To circumvent compromised routers, prior research proposes to incorporate trust into router selection. The first work is proposed by Puttaswamy et al. [1]. It allows users to select onion routers from their friends or friends of friends in online social networks [8]. Drac system [3] uses a similar technique, but it is mainly designed to facilitate anonymous communication against a global passive adversary in a peer-to-peer fashion. The first general trust model for onion routing is proposed by Johnson and Syverson [2] in 2009. This model reasons about the trust as the difficulty of compromising onion routers, but ignore the fact that different users may trust different parts of the network. To address this issue, Johnson et al. [4] presents a more comprehensive trust-based onion routing in 2011. This model considers users with different trust distribution in the network. Moreover, Marques et al. [25] report a preliminary survey for trust models used in anonymous communication.

Although trust models help evade compromised routers, the adversary who has the knowledge of a-priori trust relationships is more likely to deanonymize users by making inference. Diaz et al. [21] present a pioneer research to discuss this attack. It assumes the source and destination of a communication in a mix-based network [26, 27] are also members of a social network. The adversary who obtains the social network graph in advance can reduce the anonymity protected by the mix-based network. Johnson et al. [4] also discuss a similar attack in trust-based onion routing. They propose a downhill algorithm to mitigate the adversary's inferences. This algorithm considers the fact that the compromised routers in a user's connection close to this user are more effective in compromising anonymity than the routers far away from this user, and therefore suggests a routing algorithm that allows the user to select routers from sets with a decreasing minimum trust threshold. This algorithm does not leverage trust degree information in the design space, thus loosing the chance to further improve anonymity by selecting onion routers that are trusted by more other users.

The trust we consider in this paper is very different from other two notions of trust. One is the behavioral trust that represents the performance reputation [28–32], and the other is the environmental trust that defines the security of software and hardware platforms where the anonymity toolkits run [33].

3 Preliminaries

In this section, we first present the general trust model in trust-based onion routing. We then elaborate on the adversary model. After that, we formalize the trust degree. We also give a brief description of the anonymity protected by onion routing networks.

3.1 The Trust Model

We consider the general trust model proposed by Johnson et al. [4]. It provides a foundation for trust-based onion routing in several aspects. First, this model reasons about trust for the onion routing protocol and describes the notion of trust as the difficulty of compromising the onion routers. This difficulty represents the probability that the adversary is failed to compromise the routers. Second, this model considers a very coarse level of trust in onion routers. It is a reasonable consideration because users need outside knowledge to estimate the trust. This includes the knowledge of the technical competence of individuals who operate the routers, the computer platform where the routers are running in, and the likelihood that the router is deployed by the adversary, etc. Therefore, it is unrealistic to expect an accurate trust assigned to the routers. Third, since different users have different adversaries, this model investigates different users with different distributions of trust in routers. For example, organizations may deploy onion routers to serve their own members but attack the users from their rival.

In this model, V is the set of nodes in a trust-based onion routing network. $V = U \bigcup R \bigcup \Delta$, where U is the set of users (e.g., the human beings or their computers), R is the set of onion routers, and Δ is the set of the destinations (e.g., the web servers). c_{ij} is the probability that the onion router $r_j \in R$ is successfully compromised by u_i's adversaries. $C = [c_{ij}]^{|U| \times |R|}$ is the matrix of the probabilities for each user's adversaries compromising each router in the network. $|U|$ and $|R|$ are the number of users and onion routers in the network, respectively. $T = [t_{ij}]^{|U| \times |R|} = [1 - c_{ij}]^{|U| \times |R|} = I - C$ is the matrix of users' trust distributions over routers. $t_{ij} = 1 - c_{ij}$ is the trust u_i assigns to r_j. Since this model only takes coarse level of trust into account, there are a very limited number ν of distinct values of trust in T. Such as in [2, 4], only $\nu = 2$ and $\nu = 3$ have been studied.

We use the terms "path" and "connection" interchangeably in the rest of the paper to represent an onion route consisting of several onion routers. We regard a position of a connection as a hop of this connection. To establish a connection, a user should select onion routers to fill in all the hops of its connection. In trust-based onion routing, a user makes a connection with several hops and actively selects onion routers according to T for these hops. $P = [p_{ij}]^{|U| \times |R|}$ is the matrix of probabilities that users use to select routers based on T. [1]

3.2 The Adversary

We consider two kinds of adversary in this paper. The first kind attempts to compromise onion routers in the network. If some routers in an user's connection are compromised, especially if the routers in both the first and last hops are compromised, various attacks [8–20] can be launched to deanonymize the user. The adversary could manipulate onion routers by two means. One is to compromise legal routers that already exist in the network, and the other is to deploy its own malicious routers in the network. In some worse conditions, the adversary could compromise a significant fraction of the network, such as launching the Sybil attack [12]. The trust-based onion routing algorithms are originally proposed to defend this kind of adversary. With the help of users' own trust in onion routers, they identify and exclude compromised routers in their connections.

Although the trust model can defend against the adversary who compromises onion routers, a new kind of adversary appears and poses a significant threat to trust-based onion routing [4]. This adversary deanonymizes users by making inference based on a

[1] $P = [p_{ij}]^{|U| \times |R|}$ may be different when users select routers for different positions of their connections. This will be elaborated on in Section 5.

priori trust relationships. In particular, this adversary could exploit compromised routers or malicious destinations (e.g., malicious web servers) to observe routers in connections, and then infer the original user of the connection according to the fact that users prefer to choosing their trusted routers in trust-based onion routing. In this paper, we assume that the adversary can only employ compromised routers to observe the routers in adjacent positions of the connections (i.e., adjacent hops), or use the malicious destinations to observe the router in the last hop. To face this adversary, the user runs a high risk to be deanonymized if she selects a router barely trusted by other users.

Prior research [4,21] shows it is feasible for an adversary to make inference in practice, although this adversary is required to know users' trust distributions over onion routers in advance. In realistic environment, the adversary could estimate these trust distributions through outside knowledge [2,4]. For example, the users belonging to an organization may be more likely to trust the routers deployed by this organization. In particular, if both users and routers' owners are members of social networks, the adversary can profile the trust relationships by crawling and deanonymizing online social networks [34,35]. Moreover, since the trust-based onion routing algorithm may be set up by default in softwares and shared in the public, the adversary who knows the trust distributions can also infer users' router selection probabilities [2].

In this paper, we focus on defending against the adversary who makes inference to deanonymize the user without sacrificing the capability of defending against the adversary who attempts to compromise onion routers.

3.3 The Trust Degree

Existing trust-based onion routing networks employ users' own trust to improve anonymity by thwarting the adversary who attempts to compromise routers [4], but do not consider the trust from other users. However, if the adversary deanonymizes the user by making inference based on the knowledge of a-priori trust distributions, the trust from other users plays a very important role in protecting anonymity.

We define a router's trust degree with respect to a user as the sum of other users' trust in this router. Let d_{ij} be the trust degree of the router $r_j \in R$ with respect to the user u_i as:

$$d_{ij} = \sum_{u_x \in U} t_{xj} - t_{ij} = \sum_{u_x \in U/u_i} t_{xj} \quad (1)$$

where t_{ij} is the trust u_i assigns to r_j, t_{xj} is the trust $u_x \in U/u_i$ assigns to r_j and U/u_i is the set of users excluding u_i.

As elaborated on in Section 3.2, the adversary can estimate the trust-based router selection distributions if they have the knowledge of a-priori trust relationships and the corresponding trust-based router selection strategies. However, a user's router selection distribution may not be the same as this user's trust distribution over routers. For example, according to the trust-based algorithms proposed by Johnson and Syverson [2], if the adversary compromises a significant fraction of the network, u_i should choose the most trusted routers with the probability 1 rather than $\frac{\max_{r_j \in R} t_{ij}}{\sum_{r_j \in R} t_{ij}}$ to maximize the capability of keeping from compromised routers. The adversary could infer the user with higher accuracy by using the router selection distributions rather than the trust distributions. Therefore, a more accurate definition of a router's trust degree with respect to a user could be the sum of other users' selection probabilities for this router:

$$d_{ij} = \sum_{u_x \in U} p_{xj} - p_{ij} = \sum_{u_x \in U/u_i} p_{xj} \tag{2}$$

where p_{ij} is the probability that u_i uses to select r_j and p_{xj} is the probability that $u_x \in U/u_i$ uses to select r_j. In the rest of the paper, we use Eqn.(2) to calculate d_{ij}.

3.4 The Anonymity

The onion routing protocol keeps the adversary from linking the source and destination of a connection that a user [2] makes, hence protects the anonymity of who is talking to whom in a communication [6]. As a result, the path anonymity of a connection can be protected if the user or the destination of this connection can be concealed. When the source link (i.e., the user) of a connection can be observed by the adversary, the path anonymity depends on the destination's anonymity. In this case, Johnson et al. [4] conclude that the path anonymity can be best protected if users select one of their most trusted routers to make a single hop connection.

But if the destination link of a connection can be observed, the protection of path anonymity relies on the protection of the user anonymity. This is a common scenario in the real world. For example, an organization imposes censorship on some sensitive web sites and attempts to record who access these sites. In this paper, we focus on the problem of protecting the user anonymity when the destination link can be observed.

4 Trust Degree in Router Selection

In existing trust-based onion routing networks, users select routers only according to their own trust, thus being vulnerable to the adversary who makes inference based on a-priori trust relationships [4]. However, we find that the routers trusted by more other users are more effective in defending against this inference. Therefore, we incorporate the trust from other users into trust-based onion routing.

In this section, we elaborate on selecting routers for a single hop based on trust degree information. Section 4.1 defines the metric of anonymity for router selection. In particular, we use the chance of a user to be inferred by the adversary to measure anonymity. Section 4.2 presents the optimal router selection strategy by considering routers' trust degree to maximize anonymity. We also analyze the anonymity improvement with the help of the optimal strategy in both simulation and real-world social networks in Section 4.3. This is compared with existing trust-based strategy. Table 1 summarizes important notations used in this section.

[2] In this paper, a user actively selects routers to initiate a connection and access a destination through this connection.

Table 1. Important notations in Section 4.

Symbol	Definition	Symbol	Definition
$\|A\|$	the size of set A	$[a_{ij}]^{I\times J}$	an $I \times J$ matrix of elements a_{ij}
t_{ij}	u_i's trust in router r_j	d_{ij}	r_j's trust degree with respect to u_i , $d_{ij} = \sum_{u_i \in U \setminus u_i} p_{ij}$
p_{ij}	u_i's probability to select router r_j	R_e	a set of routers that u_i equally trusts, $\forall r_j \in R_e, t_{ij} = t_e$
$U \setminus u_i$	the set of users excluding u_i	$p_i\{R_e\}$	u_i's strategy to select routers from R_e, $p_i\{R_e\} = [p_{ij}]^{1\times\|R_e\|}$
		$\Gamma\left(p_i\{R_e\}\right)$	the expectation of the chance to infer u_i for strategy $p_i\{R_e\}$
D_e	$D_e = \sum_{r_j \in R_e} d_{ij}$	θ_e	$\theta_e = \sum_{r_j \in R_e} p_{ij}$

4.1 Minimizing the Chance of Being Inferred in Router Selection

We investigate a user u_i who is aware of routers' trust degree with respect to a population of other users whose trust distributions and router selection strategies are known. To preserve the capability of defending against compromised routers, we only consider the trust degree information for the routers equally trusted by u_i. The number of routers equally trusted by u_i could be large due to the small number of distinct trust levels considered in existing trust-based onion routing [2, 4]. Moreover, as a person's friends always receive different amount of trust from other persons [22], the routers equally trusted by u_i are more likely to have different trust degrees.

We consider the scenario that the adversary makes inference according to the observation on a single hop of u_i's connection. It may be the case that the adversary manipulates the destination and observes the last hop (i.e., the Hop-X in Figure 3).

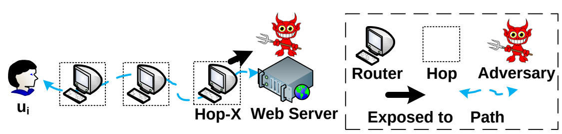


Fig. 3. An example of the single hop that can be observed.

Since the adversary has the knowledge of a-priori trust relationships and users' router selection strategies in the network, she gets the probability $\frac{p_{ij}}{p_{ij}+d_{ij}}$ to infer u_i if the router r_j is observed, where d_{ij} can be calculated by Eqn.(2). Moreover, if u_i has the probability p_{ij} to choose r_j for the exposed hop, the adversary has the probability p_{ij} to observe r_j in this hop of u_i's connection. Therefore, u_i has the probability $p_{ij} \cdot \frac{p_{ij}}{p_{ij}+d_{ij}}$ to be inferred through r_j in the exposed hop.

We consider the problem of minimizing u_i's chance of being inferred when a single hop of u_i's path is observed by the adversary. The objective function is defined as:

$$\Gamma\left(p_i\{R_e\}\right) = \sum_{r_j \in R_e} p_{ij} \cdot \frac{p_{ij}}{p_{ij}+d_{ij}} \tag{3}$$

where, $R_e \subseteq R$ is a set of routers to which the user u_i assigns the equal trust t_e, i.e., $\forall r_j \in R_e$, $t_{ij} = t_e$. R is the set of onion routers in the entire network. $p_i\{R_e\} = [p_{ij}]^{1\times|R_e|}$ is a selection strategy that u_i uses to select a router from R_e for the exposed hop. Herein, p_{ij} is a probability for u_i to select router r_j, the matrix $[p_{ij}]^{1\times|R_e|}$ consists all the p_{ij}s for $r_j \in R_e$ and $|R_e|$ is the size of set R_e.

The objective function $\Gamma(p_i\{R_e\})$ calculates the expectation of the chance to be inferred when u_i uses strategy $p_i\{R_e\}$ to select routers from R_e. A lower chance of being inferred means more anonymity for u_i. To maximize u_i's anonymity, we should find the optimal strategy $p_i{}^*\{R_e\}$ to minimize $\Gamma(p_i\{R_e\})$. We formalize this as:

$$p_i{}^*\{R_e\} = \arg\min_{p_i\{R_e\}} \Gamma(p_i\{R_e\}), \quad \mathit{subject\ to} \sum_{r_j \in R_e} p_{ij} = \theta_e \tag{4}$$

where, $\theta_e = \sum\limits_{r_j \in R_e} p_{ij}$ is the sum of u_i's probabilities of choosing routers from R_e. Existing trust-based algorithms decide θ_e. For example, If u_i is only allowed to select the most trusted routers, $\theta_e = 1$ for R_e with $t_e = \max_{r_j \in R} t_{ij}$ and $\theta_e = 0$ for other R_e. Since a user's trust in a router is modeled as the difficulty of this user's adversary in compromising this router [2, 4], the routers with equal trust from a user should have the same probability of being not compromised by this user's adversary. Therefore, to preserve the capability of defending against compromised routers, we should not change the value of θ_e when we minimize $\Gamma(p_i\{R_e\})$.

4.2 The Optimal Router Selection Strategy

As existing trust-based algorithms do not consider routers' trust degree, u_i can only use an equal probability to select routers with equal trust (i.e., $p_{ij}^{=} = \frac{\theta_e}{|R_e|}$ for $r_j \in R_e$) [2, 4]. However, by considering the trust from other users, u_i can intuitively gain more anonymity by using a higher probability to select routers trusted by more other users.

Let $[d_{ij}]^{1\times|R_e|}$ be the matrix of d_{ij} for $r_j \in R_e$. Let $D_e = \sum\limits_{r_j \in R_e} d_{ij}$ be the sum of trust degree d_{ij} for $r_j \in R_e$.

Considering $[d_{ij}]^{1\times|R_e|}$, we prove an optimal router selection strategy for u_i to minimize the chance of being inferred. Lemma 1 gives this optimal solution $p_i{}^*\{R_e\}$ and shows the minimal chance of being inferred $\Gamma(p_i^*\{R_e\})$ in theory. In $p_i{}^*\{R_e\}$, u_i's probability of choosing a router $r_j \in R_e$ is proportional to d_{ij}. The minimal chance $\Gamma(p_i^*\{R_e\})$ is inversely proportional to D_e.

Lemma 1. *Subject to* $\sum\limits_{r_j \in R_e} p_{ij} = \theta_e$*, the optimal strategy* $p_i{}^*\{R_e\}$ *for minimizing* $\Gamma(p_i\{R_e\})$ *is* $p_i{}^*\{R_e\} = [p_{ij}^*]^{1\times|R_e|} = \frac{\theta_e}{D_e} \cdot [d_{ij}]^{1\times|R_e|}$*. The minimum chance is* $\Gamma(p_i^*\{R_e\}) = \sum\limits_{r_j \in R_e} p_{ij}^* \cdot \frac{p_{ij}^*}{d_j} = \frac{(\theta_e)^2}{\theta_e + D_e}$.

Proof. In R_e, we have $|R_e|$ routers denoted as $r_1, r_2, \cdots, r_{|R_e|}$. We assume the sum of probability that u_i uses to choose r_1 and r_2 is $\beta \le \theta_e$. We first consider the problem of finding the optimal strategy for u_i to select r_1 and r_2 and minimize $p_{i1} \cdot \frac{p_{i1}}{p_{i1}+d_{i1}} + p_{i2} \cdot \frac{p_{i2}}{p_{i2}+d_{i2}}$. This problem can be formalized as below:

$$p_i^*\{r_1, r_2\} = \arg\min_{p_i\{r_1,r_2\}} (p_{i1} \cdot \tfrac{p_{i1}}{d_{i1}+p_{i1}} + p_{i2} \cdot \tfrac{p_{i2}}{d_{i2}+p_{i2}}), \quad s.t.,\ p_{i1} + p_{i2} = \beta \le \theta_e$$

As $p_{i2} = \beta - p_{i1}$, $\min\limits_{p_i(r_1,r_2)} (p_{i1} \cdot \frac{p_{i1}}{d_{i1}+p_{i1}} + p_{i2} \cdot \frac{p_{i2}}{d_{i2}+p_{i2}})$ can be written as $\min\limits_{p_{i1} \in [0,\beta]} f(p_{i1})$, where, $f(p_{i1}) = p_{i1} \cdot \frac{p_{i1}}{d_{i1}+p_{i1}} + (\beta - p_{i1}) \cdot \frac{(\beta - p_{i1})}{d_{i2}+\beta-p_{i1}}$. We know that, if $f(p_{i1})$'s second derivative is larger than 0, $f(p_{i1})$ has a minimum value. And this minimum value can be obtained when $f(p_{i1})$'s first derivative equals to 0. Such that, if $f''(p_{i1}) = \frac{d^2 f(p_{i1})}{d^2 p_{i1}} > 0$, $f(p_{i1})$ reach its minimum when $f'(p_{i1}) = \frac{df(p_{i1})}{dp_{i1}} = 0$. As $\beta \geq p_{i1} \geq 0$ and $d_{i1} > 0, d_{i2} > 0$, then we have:

$$f''(p_{i1}) = 2d_{i2}^2 \cdot p_{i1} + 2d_{i1}^2(\beta - p_{i1}) + 2d_{i1}(d_{i1}d_{i2} + d_{i2}^2) > 0.$$

Therefore, $f(p_{i1})$ has a minimum value when $f'(p_{i1}) = 0$, such as:

$$f'(p_{i1}) = (d_{i2}^2 - d_{i1}^2) \cdot p_{i1}^2 + 2d_{i1}(d_{i1}d_{i2} + d_{i1}\beta + d_{i2}^2) \cdot p_{i1} - d_{i1}^2\beta(2d_{i2} + \beta) = 0$$

By solving this quadratic equation, we can get two roots. But considering $p_{i1} \geq 0$, we only use the positive result $p_{i1} = \frac{d_{i1}}{d_{i1}+d_{i2}} \cdot \beta$. We thus have:

$$p_{i1}^* = \frac{d_{i1}}{d_{i1}+d_{i2}} \cdot \beta, \quad p_{i2}^* = \beta - p_{i1} = \frac{d_{i2}}{d_{i1}+d_{i2}} \cdot \beta$$

and the minimum value of $(p_{i1} \cdot \frac{p_{i1}}{d_{i1}+p_{i1}} + p_{i2} \cdot \frac{p_{i2}}{d_{i2}+p_{i2}})$ is:

$$\min_{p_i(r_1,r_2)} (p_{i1} \cdot \frac{p_{i1}}{d_{i1}+p_{i1}} + p_{i2} \cdot \frac{p_{i2}}{d_{i2}+p_{i2}}) = p_{i1}^* \cdot \frac{p_{i1}^*}{d_{i1}+p_{i1}^*} + p_{i2}^* \cdot \frac{p_{i2}^*}{d_{i2}+p_{i2}^*} = \frac{\beta^2}{d_{i1}+d_{i2}+\beta}$$

Based on that, we have:

$$(\frac{p_{i1}^2}{d_{i1}+p_{i1}} + \frac{p_{i2}^2}{d_{i2}+p_{i2}}) \geqslant \frac{\beta^2}{d_{i1}+d_{i2}+\beta} = \frac{(p_{i1}+p_{i2})^2}{d_{i1}+d_{i2}+(p_{i1}+p_{i2})}$$

and when $p_{i1} = \frac{d_{i1}}{d_{i1}+d_{i2}} \cdot \beta, \quad p_{i2} = \frac{d_{i2}}{d_{i1}+d_{i2}} \cdot \beta$, the equality satisfies.

Subject to $\sum_{r_j \in R_e} p_{ij} = \theta_e$, we minimize $\Gamma(p_i\{R_e\})$ using above inequation as:

$$\begin{aligned}
\Gamma(p_i\{R_e\}) &= \sum_{j=1}^{|R_e|} p_{ij} \cdot \frac{p_{ij}}{d_{ij}+p_{ij}} = (\frac{p_{i1}^2}{d_{i1}+p_{i1}} + \frac{p_{i2}^2}{d_{i2}+p_{i2}}) + \sum_{j=3}^{|R_e|} p_{ij} \cdot \frac{p_{ij}}{d_{ij}+p_{ij}} \\
&\geq \frac{(p_{i1}+p_{i2})^2}{d_{i1}+d_{i2}+(p_{i1}+p_{i2})} + \frac{p_{i3}^2}{d_{i3}+p_{i3}} + \sum_{j=4}^{|R_e|} p_{ij} \cdot \frac{p_{ij}}{d_{ij}+p_{ij}} \\
&\geq \frac{(p_{i1}+p_{i2}+p_{i3})^2}{d_{i1}+d_{i2}+d_{i3}+(p_{i1}+p_{i2}+p_{i3})} + \frac{p_{i4}^2}{d_{i4}+p_{i4}} + \sum_{j=5}^{|R_e|} p_{ij} \cdot \frac{p_{ij}}{d_{ij}+p_{ij}} \\
&\geq \cdots \geq \frac{(\sum_{r_j \in R_e} p_{ij})^2}{\sum_{r_j \in R_e} p_{ij} + \sum_{r_j \in R_e} d_{ij}} = \frac{(\theta_e)^2}{\theta_e + \sum_{r_j \in R_e} d_{ij}} = \frac{(\theta_e)^2}{\theta_e + D_e}
\end{aligned}$$

When $p_{ij} = \frac{d_{ij}}{D_e} \cdot \theta_e$, all the equalities satisfy.

Therefore, we have the optimal strategy $p_i{}^*\{R_e\} = [p_{ij}^*]^{1\times|R_e|} = \frac{\theta_e}{D_e} \cdot [d_{ij}]^{1\times|R_e|}$ to minimize $\Gamma(p_i\{R_e\})$, i.e., $\min_{p_i\{R_e\}} \Gamma(p_i\{R_e\}) = \Gamma(p_i^*\{R_e\}) = \sum_{r_j \in R_e} p_{ij}^* \cdot \frac{p_{ij}^*}{p_{ij}^*+d_{ij}} = \frac{(\theta_e)^2}{\theta_e+D_e}$. Lemma 1 is proved. □

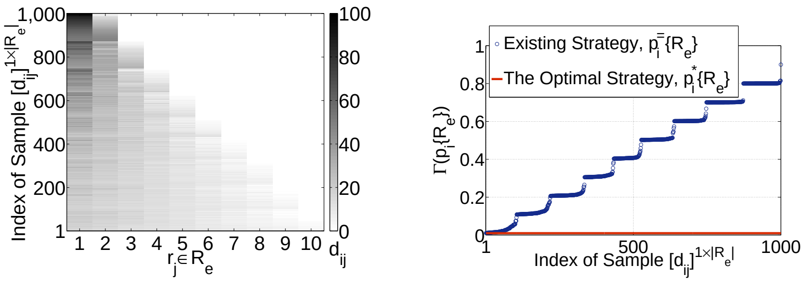
4.3 More Anonymity Through Trust Degree

We demonstrate that u_i can gain more anonymity by considering routers' trust degree in both simulation and real-world social networks. This is compared with the strategy used by existing trust-based algorithms, where the equal probability is used to select routers with equal trust [4]. We denote this existing trust-based strategy as $p_i^{=}\{R_e\} = [p_{ij}^{=}]^{1\times|R_e|}$, where $p_{ij}^{=} = \frac{\theta_e}{|R_e|}$ for $\forall r_j \in R_e$. Although the optimal strategy $p_i^*\{R_e\}$ is proved to be able to maximize u_i's anonymity, we show that u_i can gain different anonymity improvement in the context of different $[d_{ij}]^{1\times|R_e|}$. We use $\Gamma\left(p_i\{R_e\}\right)$ as the metric for u_i's anonymity. A smaller $\Gamma\left(p_i\{R_e\}\right)$ represents more anonymity. As θ_e will not affect our analysis, we simply consider $\theta_e = 1$.

Simulation We consider the case that u_i has 10 equally trusted routers (i.e., $|R_e| = 10$) and the sum of d_{ij} for $r_j \in R_e$ is 100 (i.e., $D_e = 100$). Figure 4(a) shows the heat map for 1000 different samples of $[d_{ij}]^{1\times|R_e|}$ that we randomly generate. The dark color indicates a large d_{ij} while the light color means a small value. Figure 4(b) illustrates the comparison of u_i's anonymity for these 1000 samples of $[d_{ij}]^{1\times|R_e|}$ between existing trust-based strategy (i.e., $p_i^{=}\{R_e\}$) and the optimal strategy (i.e., $p_i^*\{R_e\}$). In Figure 4(a), we sort the indexes of the 1000 samples of $[d_{ij}]^{1\times|R_e|}$ in an ascending order according to $\Gamma\left(p_i^{=}\{R_e\}\right)$ of these samples and arrange d_{ij}s in each $[d_{ij}]^{1\times|R_e|}$ in a descending order according to r_j.

Figure 4(b) shows that the $\Gamma\left(p_i^*\{R_e\}\right)$ stays at 0.0099 for any $[d_{ij}]^{1\times|R_e|}$. The value 0.0099 is the minimal chance of inferring u_i when $D_e = 100$ and $\theta_e = 1$ because $\frac{(\theta_e)^2}{\theta_e+D_e} = \frac{1}{101} = 0.0099$. Refer to Figure 4(a), we find that, a larger anonymity improvement (i.e., a larger $\frac{\Gamma(p_i^{=}\{R_e\})}{\Gamma\left(p_i^*\{R_e\}\right)}$) could be achieved in the context of $[d_{ij}]^{1\times|R_e|}$ whose d_{ij}s vary more significantly. In particular, when $[d_{ij}]^{1\times|R_e|}$ satisfies $\exists d_{ij} = 100$ and other d_{ij}s are all equal to 0, the $\Gamma\left(p_i\{R_e\}\right)$ is reduced from 0.9001 in $p_i^{=}\{R_e\}$ to 0.0099 in $p_i^*\{R_e\}$. The value 0.9001 indicates u_i suffers more than 90% probability to be inferred while 0.0099 represents this probability is less than 1%. Even when $[d_{ij}]^{1\times|R_e|}$ are uniformly distributed, i.e., d_{ij}s for $\forall r_j \in R_e$ are all the same and equal to $\frac{D_e}{|R_e|} = 10$, the optimal strategy can at least keep anonymity the same as in existing strategy (i.e., $\frac{\Gamma(p_i^{=}\{R_e\})}{\Gamma\left(p_i^*\{R_e\}\right)} = 1$).

Real-world Social Networks We also investigate the optimal strategy $p_i^*\{R_e\}$'s effectiveness by using the public data set from the Facebook [22]. This set includes more than 1.5 millions social links from 53,609 persons to their friends. Each person has more than one friend and all these 53,609 persons have 63,406 friends in total. We thus regard the 53,609 persons as the users in onion routing networks and assume the 63,406 friends deploy onion routers. We consider all these 53,609 persons as to be u_i one by one, and compare u_i's anonymity between the optimal strategy $p_i^*\{R_e\}$ and existing trust-based strategy $p_i^{=}\{R_e\}$. Each person equally trusts the routers set up by their friends, but distrusts other routers (i.e., two levels of trust are considered). Persons only select routers from their friends (i.e., $\theta_e = 1$ for R_e where $t_e = \max_{r_j\in R} t_{ij}$). We measure u_i's anonymity using $\Gamma\left(p_i\{R_e\}\right)$ and a smaller $\Gamma\left(p_i\{R_e\}\right)$ indicates more



(a) Heat map of 1000 random samples. (b) Anonymity comparison.

Fig. 4. Anonymity comparison between existing trust-based strategy and the optimal strategy for 1000 random samples of $[d_{ij}]^{1\times|R_e|}$ when $|R_e| = 10$ and $D_e = 100$.

anonymity. Note that, when a person is considered as u_i, we calculate d_{ij} for this person in the case that other persons use existing trust-based strategy to choose routers.

Figure 5 shows the results for these $53,609$ users. The D_es of these users are from 0.01 to 2491. In accordance with Lemma 1, although $\Gamma\left(p_i^*\{R_e\}\right)$ decreases when D_e increases, $\Gamma\left(p_i^*\{R_e\}\right)$ is consistently smaller than $\Gamma\left(p_i^{=}\{R_e\}\right)$ for any D_e. By analyzing the results in depth, we find more than 99.6% users can improve their anonymity with the help of the optimal strategy $p_i^*\{R_e\}$ (i.e., $\frac{\Gamma(p_i^{=}\{R_e\})}{\Gamma\left(p_i^*\{R_e\}\right)} > 1$). In particular, more than 65.6% users obtain at least 1.5 times improvement for their anonymity (i.e., $\frac{\Gamma(p_i^{=}\{R_e\})}{\Gamma\left(p_i^*\{R_e\}\right)} > 1.5$). The largest improvement is $\frac{\Gamma(p_i^{=}\{R_e\})}{\Gamma\left(p_i^*\{R_e\}\right)} = 31.1$. It can be seen that the user anonymity can be improved by considering routers' trust degree in practice.

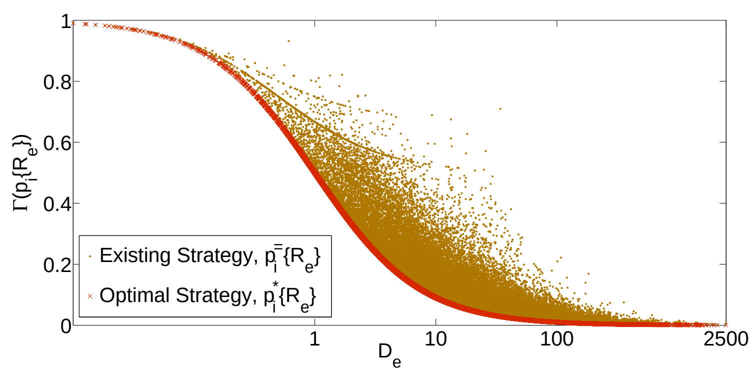


Fig. 5. Anonymity comparison between existing trust-based strategy and the optimal strategy in real-world social networks [22].

5 Trust Degree Aware Routing Algorithm for Path Selection

The scenario discussed in Section 4 assumes the adversary only observes a single hop of a connection. However, a more common scenario is that the adversary can observe more than one hop in a connection. By taking this general case into account, we design trust degree aware routing algorithms for path selection. We still only consider trust degree information among the routers equally trusted by a user. This helps preserve the capability of circumventing compromised routers.

Section 5.1 first formalizes the metric of anonymity for path selection. In particular, we measure anonymity by using the chance of a user to be inferred by the adversary who observes multiple hops of this user's connection. Section 5.2 then gives a general version of the optimal trust degree aware routing algorithm for path selection in theory.

Table 2 summarizes important notations used in this section.

Table 2. Important notations in Section 5.

Symbol	Definition
$A \setminus B$	the set A excluding a sub set $B \subseteq A$ or an element $B \in A$
h_k	the k-th hop in u_i's path
O	the set of hops exposed to the adversary
o_n	the n-th element of set O
t_k	a trust threshold for u_i to select routers in hop h_k
p_{ij}^k	u_i's probability to select router r_j for hop h_k
R_+^n	a set of routers where $r_j \in R_+^n, t_{ij} \geq t_n$
R_e^n	a set of routers with equal trust from u_i, $r_j \in R_e^n, t_{ij} = t_e \geq t_k$
$p_i\{R_+^k\}\|_O$	a routing algorithm $p_i\{R_+^k\}\|_O = \{p_i\{R_+^k\}, h_k \in O\}$
$p_i\{R_e^k\}\|_O$	a sub routing algorithm $p_i\{R_e^k\}\|_O = \{p_i\{R_e^k\}, h_k \in O\}$
N	$N = \|O\|$ be the number of exposed hops
$\Gamma\left(p_i\{R_+^k\}\|_O\right)$	the expectation of the chance to infer u_i if $p_i\{R_+^k\}\|_O$ is used
θ_e^k	$\theta_e^k = \sum_{r_j \in R_e^k} p_{ij}^k$
$D_e^{(n)} \cdot d_{ij}^{(n+1,N)}$	$\sum_{r_j \in R_e^k, h_k = o_n} \cdots \sum_{r_j \in R_e^k, h_k = o_1} \sum_{u_x \in U \setminus u_i} \prod_{h_k \in O} p_{xj}^k$

5.1 Minimizing the Chance of Being Inferred when Multiple Hops Exposed

Similar to Section 4.1, we focus on a user u_i who is aware of routers' trust degree with respect to a population of other users whose trust distributions and routing algorithms are known. Given a path of u_i, the adversary attempts to compromise routers in this path and employs the compromised routers to observe routers in adjacent hops. In particular, if the destination (e.g., a web server) is compromised, the last hop can be observed by the adversary. Based on the knowledge of a-priori trust relationships, the adversary infers u_i by observing routers in exposed hops. [3]

Given a L-hop path of u_i. Let h_k be the k-th hop in the path. Let O be the set of hops exposed to the adversary. Therefore, u_i has the probability $\prod_{h_k \in O} p_{ij}^k \cdot \prod_{h_k \in O} p_{ij}^k / \sum_{u_x \in U} \prod_{h_k \in O} p_{xj}^k$ to be inferred through r_j in each of these exposed hops, where p_{ij}^k is the u_i's probability to select r_j for the k-th hop (i.e., hop h_k) in u_i's path.

Let $o_n \in O$ be the n-th element of the set O. Let $N = |O| \leq L$ be the number of exposed hops.

[3] Prior research [4] assumes the length of users' paths is fixed and known. The adversary thus can know the number of unexposed hops in the path and make some inference based on these unexposed hops. In this paper, we do not consider the inference based on unexposed hops because the user can simply establish path with random length to evade such inference.

We consider the problem as minimizing the chance of being inferred when a set O of hops in u_i's path are observed by the adversary. The objective function is:

$$\Gamma\left(p_i\{R_+^k\}|_O\right) = \sum_{r_j \in R_+^k, h_k = o_N} \cdots \sum_{r_j \in R_+^k, h_k = o_1} \frac{\prod\limits_{h_k \in O} p_{ij}^k \cdot \prod\limits_{h_k \in O} p_{ij}^k}{\sum\limits_{u_x \in U} \prod\limits_{h_k \in O} p_{xj}^k} \tag{5}$$

where, $p_i\{R_+^k\}|_O = \{p_i\{R_+^k\}, h_k \in O\}$ is a routing algorithm consisting of $N = |O|$ router selection strategies for these exposed hops belonging to O in the path. Each $p_i\{R_+^k\} = [p_{ij}^k]^{1\times|R_+^k|}$ is a router selection strategy for the k-th hop (i.e., h_k). $R_+^k \subseteq R$ is the set of candidate routers that u_i can select for hop h_k, i.e., $\sum_{r_j \in R_+^k} p_{ij}^k = 1$. Existing trust-based routing algorithms will use a trust threshold t_k to restrict u_i's router selection for its hop h_k, such as $\forall r_j \in R_+^k, t_{ij} \geq t_k$. In particular, the downhill algorithm [4] uses a decreasing trust threshold in the hops from the user to the destination, i.e., $t_k \leq t_{k-1}$. But if u_i is only allowed to select the most trusted routers for its connection, $t_k = \max\limits_{r_j \in R} t_{ij}$ for $\forall k \in [1, L]$.

Let R_e^k be a set of routers with equal trust $t_e \geq t_k$ (i.e., $r_j \in R_e^k$, $t_{ij} = t_e \geq t_k$). We can express R_+^k as $R_+^k = \bigcup_{t_e \geq t_k} R_e^k$.

The object function $\Gamma\left(p_i\{R_+^k\}|_O\right)$ calculates the expectation of the chance that u_i can be inferred when the routing algorithm $p_i\{R_+^k\}|_O$ is used. As a lower chance of being inferred indicates more anonymity, we maximize u_i's anonymity by finding the optimal routing algorithm $p_i\{R_+^k\}|_O^*$ to minimize $\Gamma\left(p_i\{R_+^k\}|_O\right)$ as:

$$\begin{aligned} p_i\{R_+^k\}|_O^* = \arg \min_{p_i\{R_+^k\}|_O} \Gamma\left(p_i\{R_+^k\}|_O\right),\ where,\ R_+^k = \bigcup_{t_e \geq t_k} R_e^k \\ subject\ to \sum_{r_j \in R_e^k} p_{ij}^k = \theta_e^k\ for\ t_e \geq t_k\ and\ h_k \in O \end{aligned} \tag{6}$$

where, θ_e^k is the sum of u_i's probabilities of choosing routers with equal trust $t_e \geq t_k$ for hop h_k in u_i's connection. We should keep any θ_e^k the same as in existing trust-based algorithms when we explore the optimal $p_i\{R_+^k\}|_O^*$, because the same θ_e^k means the same capability of defending against compromised routers.

Let $p_i\{R_e^k\}|_O = \{p_i\{R_e^k\}, h_k \in O\}$. As $R_+^k = \bigcup_{t_e \geq t_k} R_e^k$, the object function in Eqn.(5) thus can be re-expressed as:

$$\begin{aligned} \Gamma\left(p_i\{R_+^k\}|_O\right) = \sum_{t_e \geq t_k, h_k = o_N} \cdots \sum_{t_e \geq t_k, h_k = o_1} \Gamma\left(p_i\{R_e^k\}|_O\right), \\ where,\ \Gamma\left(p_i\{R_e^k\}|_O\right) = \sum_{r_j \in R_e^k, h_k = o_N} \cdots \sum_{r_j \in R_e^k, h_k = o_1} \frac{\prod\limits_{h_k \in O} p_{ij}^k \cdot \prod\limits_{h_k \in O} p_{ij}^k}{\sum\limits_{u_x \in U} \prod\limits_{h_k \in O} p_{xj}^k} \end{aligned} \tag{7}$$

Therefore, to facilitate the exploration of the minimal $\Gamma\left(p_i\{R_+^k\}|_O\right)$ without changing the value of any θ_e^k for $t_e \geq t_k, h_k \in O$, we can find the minimal $\Gamma\left(p_i\{R_e^k\}|_O\right)$ subject to each θ_e^k instead. When all the minimal $\Gamma\left(p_i\{R_e^k\}|_O\right)$s for $t_e \geq t_k, h_k \in O$ are found, the minimal $\Gamma\left(p_i\{R_+^k\}|_O\right)$ is also reached. The optimal routing algorithm $p_i\{R_+^k\}|_O^*$ consists a set of sub optimal routing algorithms $p_i\{R_e^k\}|_O^*$ for $t_e \geq t_k, h_k \in O$.

5.2 The Optimal Trust Degree Aware Routing Algorithm in Theory

Intuitively, we expect the sub optimal routing algorithm $p_i\{R_e^k\}|_O^*$ can be implemented by applying the single hop's optimal router selection strategy $p_i^*\{R_e\}$ proposed in Section 4 to each of the $N = |O|$ exposed hops independently, i.e., $p_i^*\{R_e^k\} = p_i^*\{R_e\}$ for $h_k \in O$. However, it is not the case because the router selection strategies $p_i^*\{R_e^k\}$ for these N exposed hops are correlated. To illustrate it, we give an example in Figure 6. We assume u_i equally trusts routers r_1, r_2 and r_3. If only hop h_2 is exposed to the adversary, according to the single hop's optimal router selection strategy $p_i^*\{R_e\}$, we should have a larger probability to choose r_1 than r_2 for hop h_2, because r_1 is trusted by two other users (i.e., u_1 and u_2) but r_2 is just trusted by one (i.e., u_3). However, if hop h_3 is also exposed and r_3 is already selected for hop h_3, the adversary can deanonymize u_i directly if u_i selects r_1 for hop h_2. The reason is that, except u_i, no other users trust both r_1 and r_3 in Figure 6. In this situation, we cannot minimize the adversary's chance of inferring u_i by applying the single hops's optimal strategy to hop h_2 independently.

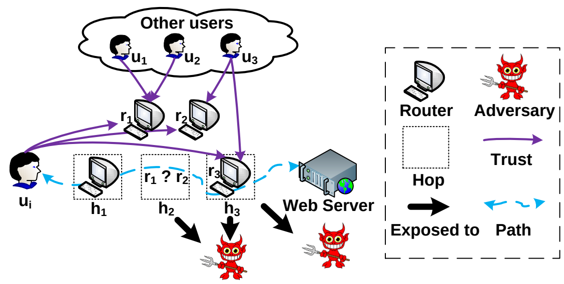


Fig. 6. An example to show the router selection strategies in different exposed hops are correlated.

Based on the analysis of Figure 6, we find that the joint probabilities of selecting routers for multiple exposed hops are correlated. We consider u_i selects routers for its connection in a descending order (i.e., given $h_k, h_{k'}$ and $k' > k$, u_i first selects routers for $h_{k'}$). In this case, to minimize the chance of being inferred, u_i's probability of selecting a router for a hop $h_k \in O$ should depend on the routers already selected in hops $h_{k'} \in O, k' > k$.

Lemma 2 gives the optimal routing algorithm $p_i\{R_e^k\}|_O^*$ and the minimal $\Gamma\left(p_i\{R_e^k\}|_O\right)$ using this algorithm. Due to the page limit, we omit the proof of Lemma 2 in this paper.

We sort O in an ascending order, i.e., for $h_k = o_n$ and $h_{k'} = o_{n+1}$, we have $k < k'$.

Let $D_e^{(n)} \cdot d_{ij}^{(n+1,N)} = \sum\limits_{r_j \in R_e^k, h_k = o_n} \cdots \sum\limits_{r_j \in R_e^k, h_k = o_1} \sum\limits_{u_x \in U \setminus u_i} \prod\limits_{h_k \in O} p_{xj}^k$, where $U \setminus u_i$ is the set of users excluding u_i. In particular, $D_e^{(0)} \cdot d_{ij}^{(1,N)} = d_{ij}^{(1,N)} = \sum\limits_{u_x \in U \setminus u_i} \prod\limits_{h_k \in O} p_{xj}^k$ and $D_e^{(N)} \cdot d_{ij}^{(N+1,N)} = D_e^{(N)} = \sum\limits_{r_j \in R_e^k, h_k = o_N} \cdots \sum\limits_{r_j \in R_e^k, h_k = o_1} \sum\limits_{u_x \in U \setminus u_i} \prod\limits_{h_k \in O} p_{xj}^k$.

Lemma 2. *Subject to* $\sum\limits_{r_j \in R_e^k} p_{ij}^k = \theta_e^k$ *for* $t_e \geq t_k, h_k \in O$*, the optimal routing algorithm* $p_i\{R_+^k\}|_O^*$ *for minimizing* $\Gamma\left(p_i\{R_+^k\}|_O\right)$ *consists of a set of sub optimal algorithms* $p_i\{R_e^k\}|_O^*$ *for* $t_e \geq t_k, h_k \in O$*. In each* $p_i\{R_e^k\}|_O^*$*, for* $h_k = o_n$*, we have:*

$$p_i^*\{R_e^k\} = [p_{ij}^{k*}]^{1\times|R_e^k|} = \frac{\theta_e^k}{D_e^{(n)} \cdot d_{ij}^{(n+1,N)}} \cdot D_e^{(n-1)} \cdot [d_{ij}^{(n,N)}]^{1\times|R_e^k|}$$

where, the hop h_k is the n-th element in O (i.e., $h_k = o_n$). Using this optimal routing algorithm, the chance can be minimized to be:

$$\min_{p_i\{R_+^k\}|_O} \Gamma\left(p_i\{R_+^k\}|_O\right) = \sum_{t_e \geq t_k, h_k = o_N} \cdots \sum_{t_e \geq t_k, h_k = o_1} \frac{(\prod_{h_k \in O} \theta_e^k)^2}{\prod_{h_k \in O} \theta_e^k + D_e^{(N)}}$$

Where, $D_e^{(n-1)} \cdot [d_{ij}^{(n,N)}]^{1\times|R_e^k|}$ is a matrix of $D_e^{(n-1)} \cdot d_{ij}^{(n,N)}$s for $r_j \in R_e^k, h_k = o_n$. Moreover, $D_e^{(n)} \cdot d_{ij}^{(n+1,N)}$ can be considered as the sum of $D_e^{(n-1)} \cdot d_{ij}^{(n,N)}$s over $r_j \in R_e^k, h_k = o_n$. Since the calculation of $D_e^{(n-1)} \cdot d_{ij}^{(n,N)}$ and $D_e^{(n)} \cdot d_{ij}^{(n+1,N)}$ are based on the p_{ij}^ks for $h_k \in \{o_{n+1}, \ldots, o_N\}$, different $r_j \in R_e^k, h_k \in \{o_{n+1}, \ldots, o_N\}$ will lead to different $p_i^*\{R_e^k\}, h_k = o_n$. In the optimal algorithm $p_i\{R_e^k\}|_O^*$, the router selection strategy $p_i^*\{R_e^k\} = [p_{ij}^{k*}]^{1\times|R_e^k|} = \frac{\theta_e^k}{D_e^{(N)}} \cdot D_e^{(N-1)} \cdot [d_{ij}^{(N,N)}]^{1\times|R_e^k|}$ for the last exposed hop $h_k = o_N$ is the base case and independent from the routers in other hops.

The optimal routing algorithm given in Lemma 2 is general and we can use it to improve any trust-based onion routing algorithms. In particular, if the trust-based algorithm restricts u_i to select its most trusted routers for its connection, the corresponding optimal trust degree aware routing algorithm is a special case of the general version when $t_k = t_e = \max_{r_j \in R} t_{ij}$ and $\theta_e^k = 1$ for $h_k \in O$. Since the downhill algorithm uses the same probability to select routers from R_+^n [4], the optimal trust degree aware downhill algorithm can be the special case of the general version when $t_k \leq t_{k-1}$ and $\theta_e^k = \frac{|R_e^k|}{|R_+^n|}$ for $t_e \geq t_k, h_k \in O$.

An Example We give an example to help understand Lemma 2 in depth. In this example, we design an optimal trust degree aware routing algorithm for u_i given the last two hops exposed (i.e., $O = \{o_1 = h_2, o_2 = h_3\}$) in Figure 6. We assume the network only includes four users (i.e., u_i, u_1, u_2 and u_3) and three onion routers (i.e., r_1, r_2 and r_3). We investigate u_i who considers trust degree information with respect to other users (i.e., u_1, u_2 and u_3) who use the equal probabilities to select routers with equal trust. We consider two levels of trust (i.e., trust and distrust) and users are restricted to select their trusted routers. u_1 and u_2 trust r_1 but distrust r_2 and r_3. u_3 equally trusts r_2 and r_3 but distrusts r_1. Therefore, we have $p_{11}^k = p_{21}^k = 1$ and $p_{32}^k = p_{33}^k = 0.5$ for $h_k \in O = \{h_2, h_3\}$. Moreover, u_i equally trusts r_1, r_2 and r_3, we have $R_+^2 = R_e^2 = R_+^3 = R_e^3 = \{r_1, r_2, r_3\}$ and $\theta_e^2 = \theta_e^3 = 1$.

If u_i uses the same probability to choose routers with equal trust for its connection (i.e., u_i's routing algorithm is $p_i\{R_+^k\}|_{\{h_2,h_3\}}^=$ where $p_{ij}^{k=} = \frac{1}{3}$ for $r_j \in R_+^k, h_k \in \{h_2, h_3\}$), the adversary has the chance $\Gamma\left(p_i\{R_+^k\}|_{\{h_2,h_3\}}^=\right) = 0.587$ to infer u_i. But if u_i uses the optimal trust degree aware routing algorithm $p_i\{R_+^k\}|_{\{h_2,h_3\}}^*$ for the 2 exposed hops according to Lemma 2, the adversary's chance of inferring u_i is minimized to $\Gamma\left(p_i\{R_+^k\}|_{\{h_2,h_3\}}^*\right) = 0.25$. It can be seen, u_i obtains more than 2 times improvement for its anonymity (i.e., $\frac{\Gamma\left(p_i\{R_+^k\}|_{\{h_2,h_3\}}^=\right)}{\Gamma\left(p_i\{R_+^k\}|_{\{h_2,h_3\}}^*\right)} > 2$). Table 3 gives this optimal algorithm. The probabilities of selecting routers for hop h_2 depend on the routers that are already selected in hop h_3.

Table 3. The optimal trust degree aware routing algorithm $p_i\{R^k_+\}|^*_{\{h_2,h_3\}}$ of u_i in Figure 6.

$r_j \in R^3_+$	r_1			r_2			r_3		
p^{3*}_{ij}	0.6667			0.1667			0.1667		
$r_j \in R^2_+$	r_1	r_2	r_3	r_1	r_2	r_3	r_1	r_2	r_3
p^{2*}_{ij}	1	0	0	0	0.5	0.5	0	0.5	0.5

6 Conclusions

In this paper, we show that the user can gain more anonymity by considering routers' trust degree in trust-based onion routing networks. With solid theoretical analysis, we propose the optimal trust degree aware solutions to maximize anonymity for both router selection and path selection. This is a theoretical foundation for trust degree aware onion routing. Our results benefit future research for practical applications.

Acknowledgements

We thank the anonymous reviewers for their very helpful comments. This work is partially supported by grants G-U386 from The Hong Kong Polytechnic University, 60903185 from NSFC, 2012M511058 from CPSF and 12R21412500 from STCSM.

References

1. Krishna P. N. Puttaswamy, Alessandra Sala, and Ben Y. Zhao. Improving anonymity using social links. In *Proc. Workshop on Secure Network Protocols*, 2008.
2. A. Johnson and P. Syverson. More anonymous onion routing through trust. In *Proc. IEEE CSF*, 2009.
3. George Danezis, Claudia Diaz, Carmela Troncoso, and Ben Laurie. Drac: An architecture for anonymous low-volume communications. In *Proc. PETS*, 2010.
4. Aaron Johnson, Paul Syverson, Roger Dingledine, and Nick Mathewson. Trust-based anonymous communication: Adversary models and routing algorithms. In *Proc. ACM CCS*, 2011.
5. David M. Goldschlag, Michael G. Reed, and Paul F. Syverson. Hiding routing information. In *Proc. Workshop on Information Hiding*, 1996.
6. Paul F. Syverson, David M. Goldschlag, and Michael G. Reed. Anonymous connections and onion routing. In *Proc. IEEE Symposium on Security and Privacy*, 1997.
7. R. Dingledine, N. Mathewson, and P. Syverson. Tor: The second-generation onion router. In *Proc. USENIX Security Symposium*, 2004.
8. Matthew Wright, Micah Adler, Brian Neil Levine, and Clay Shields. The predecessor attack: An analysis of a threat to anonymous communications systems. *ACM Transactions on Information and System Security*, 2004.
9. Nathan S. Evans, Roger Dingledine, and Christian Grothoff. A practical congestion attack on Tor using long paths. In *Proc. USENIX Security Symposium*, 2009.
10. Carmela Troncoso and George Danezis. The Bayesian traffic analysis of mix networks. In *Proc. ACM CCS*, 2009.
11. Dakshi Agrawal and Dogan Kesdogan. Measuring anonymity: the disclosure attack. *IEEE Security & Privacy*, 2003.
12. J. Douceur. The Sybil attack. In *Proc. International Workshop on Peer-To-Peer Systems*, 2002.

13. Paul Syverson, Gene Tsudik, Michael Reed, and Carl Landwehr. Towards an analysis of onion routing security. In *Proc. Designing Privacy Enhancing Technologies: Workshop on Design Issues in Anonymity and Unobservability*, 2000.
14. S. Murdoch and G. Danezis. Low-cost traffic analysis of Tor. In *Proc. IEEE Symposium on Security and Privacy*, 2005.
15. L. Øverlier and P. Syverson. Locating hidden servers. In *Proc. IEEE Symposium on Security and Privacy*, 2006.
16. K. Bauer, D. McCoy, D. Grunwald, T. Kohno, and D. Sicker. Low-resource routing attacks against Tor. In *Proc. ACM Workshop on Privacy in the Electronic Society*, 2007.
17. X. Fu and Z. Ling. One cell is enough to break Tor's anonymity. In *Proc. Black Hat DC*, 2009.
18. Z. Ling, J. Luo, W. Yu, X. Fu, D. Xuan, and W. Jia. A new cell counter based attack against Tor. In *Proc. ACM CCS*, 2009.
19. Y. Zhu, X. Fu, B. Graham, R. Bettati, and W. Zhao. Correlation-based traffic analysis attacks on anonymity networks. *IEEE Transactions on Parallel and Distributed Systems*, 2009.
20. Nicholas Hopper, Eugene Y. Vasserman, and Eric Chan-Tin. How much anonymity does network latency leak? *ACM Transactions on Information and System Security*, 2010.
21. Claudia Diaz, Carmela Troncoso, and Andrei Serjantov. On the impact of social network profiling on anonymity. In *Proc. Workshop on Privacy Enhancing Technologies*, 2008.
22. Alan Mislove. Wosn 2009 data sets. `http://socialnetworks.mpi-sws.org/data-wosn2009.html`, 2009.
23. Xiapu Luo, Peng Zhou, Junjie Zhang, Roberto Perdisci, Wenke Lee, and Rocky K. C. Chang. Exposing invisible timing-based traffic watermarks with backlit. In *Proc. ACSAC*, 2011.
24. Xiapu Luo, Junjie Zhang, Roberto Perdisci, and Wenke Lee. On the secrecy of spread-spectrum flow watermarks. In *Proc. ESORICS*, 2010.
25. Rodolphe Marques and André Zúquete. Social networking for anonymous communication systems: A survey. In *Proc. International Conference on Computational Aspects of Social Networks*, 2011.
26. David Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. *Communications of the ACM*, 1981.
27. George Danezis, Roger Dingledine, and Nick Mathewson. Mixminion: Design of a type iii anonymous remailer protocol. In *Proc. IEEE Symposium on Security and Privacy*, 2003.
28. The Tor Project. Tor path selection specification. `http://tor.hermetix.org/svn/trunk/doc/spec/path-spec.txt`, 2009.
29. R. Snader and N. Borisov. A tune-up for Tor: Improving security and performance in the Tor network. In *Proc. ISOC Network and Distributed System Security Symposium*, 2008.
30. R. Snader and N. Borisov. Improving security and performance in the Tor network through tunable path selection. *IEEE Transactions on Dependable and Secure Computing*, 2010.
31. Roger Dingledine, Michael J. Freedman, David Hopwood, and David Molna. A reputation system to increase mix-net reliability. In *Proc. International Workshop on Information Hiding*, 2001.
32. R. Dingledine and P. Syverson. Reliable MIX cascade networks through reputation. In *Proc. International Conference on Financial Cryptography*, 2003.
33. Alexander Böttcher, Bernhard Kauer, and Hermann Härtig. Trusted computing serving an anonymity service. In *Proc. International Conference on Trust & Trustworthy Computing*, 2008.
34. Ralph Gross and Alessandro Acquisti. Information revelation and privacy in online social networks. In *Proc. ACM workshop on Privacy in the electronic society*, 2005.
35. Arvind Narayanan and Vitaly Shmatikov. De-anonymizing social networks. In *Proc. IEEE Symposium on Security and Privacy*, 2009.