# Social Networking for Anonymous Communication Systems: A Survey

Rodolphe Marques
*Instituto de Telecomunicações*
*Aveiro, Portugal*
*rmarques@av.it.pt*

André Zúquete
*Universidade de Aveiro/IEETA*
*Aveiro, Portugal*
*andre.zuquete@ua.pt*

## Abstract

*Anonymous communication systems have been around for sometime, providing anonymity, enhanced privacy, and censorship circumvention. A lot has been done, since Chaum's seminal paper on mix networks, in preventing attacks able to undermine the anonymity provided by these systems. This, however, is a difficult goal to achieve due to the decentralized nature of these systems. In the end it boils down to finding a subset of trusted nodes to be placed in critical positions of the communication path. But the question remains: "How to know if a given node can be trusted?". In this paper we present a survey of a new research area which goal is to exploit trust in social links to solve some of the shortcomings of anonymous communication systems. Recent research shows that by using social networking features it is possible to prevent traffic analysis attacks and even detect Sybil attacks.*

## Index Terms

*Social Network, Security, Anonymity, Sybil Attack, Reputation*

## 1. Introduction

In recent years a new research area emerged with the goal of exploiting social networking features in order to solve some of the short comings of anonymous communication systems.

Anonymous communication systems are usually built as peer-to-peer overlay networks, that rely on a community of volunteers that are responsible for maintaining and running the nodes that constitute the network. Users then make use of this networks to remain anonymous, by creating a circuit, or multi-hop path, through a set of selected relay nodes in the network.

Volunteers are an important part of these networks, since the degree of anonymity provided by such networks depends on the number of nodes and their behavior in the network. On the other end, the open nature of these networks allow anyone willing to share some of its bandwidth to run a relay node. This leaves a door open to malicious behavior in the network, where an attacker can run one or several relay nodes in the network in an attempt to undermine the anonymity that these networks provide to their clients.

Attacks on anonymous communication systems have been thoroughly studied, most of them being passive attacks through traffic analysis. The most common attacks are the predecessor attack [1], the intersection attacks [2], timing attacks [3], and Sybil attacks [4]. These attacks can be easily performed by controlling some specific relay nodes in the circuits created by the clients. Timing attacks are more critical in low latency anonymity networks and are extremely difficult to detect do to their passive nature.

One way to thwart these attacks would be to gain some information about the nodes in the network, in order to establish trust relationships with a subset of nodes to relay clients' traffic. This model would substitute the actual model in which nodes are chosen randomly.

For the purpose of this survey we define a social network as a network of people that share a common goal and are connected through bi-directional trust relationships.

In this survey we study current proposals that try to exploit trust relationships of social links. The end goal varies, but the main idea is to find out if it is possible to create anonymity sets which are trusted, and offer protection against traffic analysis.

Anonymity sets are sets of entities that have the same probability of being the confused with the real client of an anonymization network. The larger the anonymity set, the highest the degree of anonymity provided by the set. Choosing relay nodes from a trusted anonymity set, as opposed to randomly chosen nodes, may result in smaller anonymity sets but, as we will show, this does not necessarily mean less anonymity.

This paper is organized as follows: in section 2 we give a brief overview of anonymous communication systems, in section 3 we present an overview of the most common attacks. In section 4 we present a survey of the work presented this far on how to exploit social networks to improve anonymity, followed by the conclusions in section 5.

## 2. Overview of Anonymous Communication Systems

In 1981 Chaum proposed a method to protect the identities of communicating parties, called mix networks [5]. The basic functionality of a mix network is to provide sender and receiver anonymity by using proxy relaying servers. Each message is encrypted in each proxy using public key cryptography, resulting in a layered encryption.

Onion routing [6] is based on the mix networks proposed by Chaum. It provides connections that are strongly resistant to both eavesdropping and traffic analysis. Onion Routing operates by dynamically building anonymous connections within a network of real-time Chaum mixes [5].

An onion is a recursively layered data structure that specifies properties of the connection at each point along the route. Each onion router along the route uses its public key to decrypt onions that it receives. This operation exposes the cryptographic control information, the identity of the next onion router, and the embedded onion. The onion router pads the embedded onion to maintain a fixed size, and sends it to the next onion router. After the connection is established, data can be sent in both directions. Data from the initiator is repeatedly pre-encrypted using the algorithms and keys that were specified in the onion. As data moves through the anonymous connection, each onion router removes one layer of encryption as defined by the cryptographic control information in the onion defining the route, so the data arrives as plain text at the recipient

Tor [7] is a popular software that uses onion routing. The Tor network is a distributed relay network that relays arbitrary TCP streams over a network of relay nodes called Onion Routers. Each onion router maintains a long term identity key used to sign TLS certificates and maintain a TLS connection to every other Onion Router. Each Tor client runs a local software called Onion Proxy which is responsible for fetching router lists, the current network state, establish circuits across the network, and handle connections from the user applications. Clients choose a path through the network and build a circuit, in which each node in the path knows its predecessor and successor, but no other nodes in the circuit. Traffic flows down the circuit in fixed-size cells, which are unwrapped by a symmetric key at each node and relayed downstream.

Crowds [1] prevents a Web server from learning any potentially identifying information about its clients, including the clients' IP address or domain name. Crowds works by collecting Web clients into a geographically diverse group called a "crowd" that performs Web transactions on behalf of its members. A user joins the crowd by starting a process in his local machine, called a "jondo", during which it is informed of the other current crowd members and the other crowd members are informed of the new jondo's membership. Once admitted to the crowd, it can employ the crowd to issue requests to Web servers in a way that prevents Web servers and other crowd members from determining who initiated those requests, by randomly choosing another member of the crowd to which it redirects its requests. This random request redirection is then performed several times within the crowd, until a crowd member delivers that request to the target Web server.

Freenet [8] can be seen as a social network where users can anonymously share files, browse and publish Web sites accessible only through Freenet, and chat on forums. Freenet is decentralized, in order to make it less vulnerable to attacks, and it allows an operation mode where users only connect to their friends. Freenet stores documents and allows them to be retrieved later by an associated key. The system has no central servers and is not subject to the control of any individual or organization. Freenet is implemented as a peer-to-peer network of nodes that query one another to store and retrieve files, which are named by location-independent keys.

Anonymizer [9] is a tool that attempts to make activity on the Internet untraceable by accessing the Internet on the users' behalf, hiding computer identifiers of source users. Anonymizer acts as a trusted intermediary between its users and the Web sites they are viewing, hiding the user IP address, making it almost impossible for third parties to track the sites that the user visits and build profiles of the user activities on the Internet. It is also able to protect personal information by redirecting users' traffic through its secure servers.

## 3. Common Attacks in Anonymous communication systems

The most common attacks know to anonymous communication systems are:

- **Predecessor attack:** In the predecessor attack, the attacker tracks an identifiable stream of communications over a number of rounds (path reformation). In each round, the attacker simply logs any node that sends a message that is part of the tracked stream. The attack does not always require analysis of timing or size of packets, but instead exploits the process of path initialization.
- **Intersection attack:** An attacker having information about which users are active at any given time can, through repeated observations, determine which users communicate with each other. This attack is based on the assumption that users typically communicate with a relatively small number of parties.
- **Timing attack:** In a timing attack, the attacker studies the timing of messages moving through the system to find correlations. It may be used by two or more attackers to determine that they are in the same communication path to identify a stream of communications.

The predecessor attack was first presented by Reiter et al. in the original Crowds paper [1]. The authors describe an attack that allows an attacker to guess the initiator of an anonymous connection based on the information about the predecessor on the path of proxies.

In [10] Wright et al. proved that when a particular initiator continues communications with a particular responder across path reformations, existing anonymity protocols are subject to attack. In the predecessor attack, the attacker tracks an identifiable stream of communications over a number of path reformations. In each path reformation, the attacker simply logs any node that sends a message that is part of the tracked stream.

The predecessor attack is a passive attack, and thus, it is extremely hard to detect. Timing attacks may be used with the predecessor attack in order to identify a stream of communications. In conclusion the authors proved that as long as attackers are selected uniformly at random to be part of the active set, and sessions can be identified across path reformations, the degree of anonymity of any sender will degrade under this attack.

In their follow up paper [11] Wright et al. describe a possible defense that comes from breaking the assumptions of uniformly random path selection. They studied the effects of non-uniform selection of nodes during path selection. In [12] Syverson et al. had also realized that the predecessor attack would fail, with the onion routing protocol, when the first node in the path was a trusted node.

Wright et al. [11] perform a deep analysis on the use of fixed nodes in the path selection, for both static systems (nodes do not leave during the communication session), and dynamic systems (nodes may leave during the communication session). They showed that if a fixed node is used in the first position of the path, it protects the initiator from ever being tracked by the attacker. When the attacker runs the predecessor attack on a communication session for a given receiver, they will see the fixed node as the initiator of the communication session instead of the initiator itself. The protection is similar for when the fixed node is at the end of the path, since all communications to a given receiver are hidden from the attackers by the fixed node. Ultimately the initiator may use a fixed node for both the first and last position on the path.

The authors also showed that in a dynamic system fixing the first and the last node may not provide significantly strong security over time when nodes leave the system frequently, forcing the initiator to rebuild a communication path.

These results provide a good metric when using a social network. The system could advice the user of which nodes to use at any given time of the day based on past history of the availability of the nodes in the social network. Furthermore, users of a social network would have an incentive to be available, in order to to relay their friends' traffic and shield them from attackers.

In [3] Levine et al. focus in clarifying the threat posed to low latency mix systems by timing attacks. They show that timing attacks are a serious threat and are easy to exploit by a well placed attacker in low latency mix-based systems like onion routing, which offer no special provision to prevent these kind of attacks. In such systems, if the first and last mixes on the path are compromised, effective timing analysis may allow the attacker to link sender with receiver.

A known defense against timing attacks is to use constant rate cover traffic along the length of the entire path, in order to reduce the correlation between the first and the last mixes. The problem with this approach is that it does not work when there are dropped packets, since they provide holes in the traffic, making it easier to correlate a stream of communications. To solve this problem the authors introduce a new defense against timing analysis called defensive dropping. They show that by randomly dropping some of the cover traffic with sufficiently large frequency, the correlation between the first and last mixes will be reduced.

Besides traffic analysis attacks, there is another attack that is extremely hard to detect and prevent in distributed computing environments, the Sybil attack [4]. In the context of anonymous communication networks, a Sybil attack occurs when several nodes, under the control or supervision of the same entity, work in collusion, sharing information amongst themselves in order to overthrow the anonymity provided by the network, in order to reveal the communicating endpoints.

In [4] the author argues that it is practically impossible, in distributed computing environments with no central trusted authority, to vouch for a one-to-one correspondence between entity and identity.

## 4. Exploitation of Social Networks for Improving Network Anonymization

In this section we provide an overview of recent work on the use of social networking techniques to improve the anonymity degree provided by anonymous communication systems. Recent work focuses mainly in exploiting trust relationships provided by social networks and how they can be used in selecting trusted nodes in order to prevent traffic analysis. There is also interesting work on reputation, which could be useful for rating nodes in a network. From an information theory point of view, it is shown that knowing more information about a given node in the network does not necessarily reduce anonymity. Another important application of social networks is the detection of Sybil attacks.

### 4.1. Path Building

In [13] Puttaswamy et al. propose the use of social links in order to improve anonymity through the introduction of trusted nodes in the path. The authors focus on the creation of paths using k-hop friends in the social network as guard nodes. Assuming that the nodes in the anonymous communication system belong to a social network, each source is able to build a strong anonymous path prefaced by hops in its social network.

The social links translate to a trust relationship established between friend nodes in a social network. The source nodes construct a communication path such that friends or friends-of-a-friend are inserted into the path in order to shield the source from being observed by passive attackers.

Finally the authors evaluate the security of this approach in the presence of intersection and predecessor attacks, and present an optimized path construction algorithm based on cliques (groups of people with similar interests).

In [14] Danezis et al. present a novel architecture for anonymous low-volume communications whose trust model is based on a friend-of-a-friend architecture. In the proposed architecture communications between friends are unobservable and communications with further contacts in the network are anonymous.

The use of a friend-of-a-friend architecture translates into a smaller anonymity set, since the social network of a user is only a subset of all the nodes in the network. Although the anonymity set is smaller, an attacker has to infiltrate the social circle of the user to perform an insider attacker.

In [15] Johnson, et al. propose a model of trust, in the context of onion routing, in order to design path selection strategies that minimize the probability of an attacker of controlling the first and last node on the path, and thus protecting against traffic analysis attacks.

They break the assumption that all nodes routing traffic are equally trusted. There is usually information available for those selecting paths in the network that can affect trust, such as: who runs the nodes; what computing platforms are being used by the nodes; how long and how reliably the nodes have been running.

### 4.2. Protection Against Sybil Attacks

In [16] and the follow up paper [17], Yu et al. propose a new protocol for defending against Sybil attacks without relying on a trusted central authority. They assume a social network where the nodes are identities in the distributed system and the edges correspond to human established trust relationships on the real world. The basic idea beyond the protocol is that a malicious user can create many identities but few trust relationships, which results in a social graph where there are many connections between Sybil nodes and few connections with honest nodes.

The social graph created by Sybil nodes exhibit a small set of edges whose removal disconnects a large number of nodes, the Sybil nodes, a behavior that is not presented in social networks.

### 4.3. Information Theory

In [15] the authors pose some important questions, while analyzing the implications of choosing nodes based on trust:

- Does using trusted nodes help to build a profile or identify clients?
- Has using trusted nodes more often the disadvantage of simultaneously reducing anonymity sets?
- What are the implication of using smaller but trusted anonymity sets?

Diaz et al. develop work that help answer these questions from an information theory context. In [18] the authors evaluate the anonymity properties of an abstract anonymous communication system of users, linked via a social network, that send messages to their friends, focusing in how the uncertainty in the attacker's knowledge of user profiles affects anonymity.

The authors study how the quantity, quality, and depth of knowledge about the users relationships affect their degree of anonymity, as well has the effect that the size of the social network versus the size of the user social network has on the degree of anonymity.

In [19] Diaz et al. answer the question of if the extra information that is provided by the social network decreases the degree of anonymity. The authors use Shannon entropy as an anonymity metric, and show that the combination of user profile information with observations at the communication layer does not necessarily lead to a reduction of the attacker's uncertainty, and that it may actually lead to higher entropies.

Finally, in [20] Diaz et al. present a model to measure anonymity of users associated with a profile. The main goal of the model is to find a good compromise between anonymity and usefulness of such systems. To have many different profiles means that users are easily distinguishable, and thus less anonymous. To have few different profiles means that users are more anonymous but cannot be targeted in an individual basis, which may not be a desirable property when using social networks.

The model focus on anonymity at the data level where N users, belonging to G groups, generate R requests using a mix network at the communication level. In normal anonymous communication system with N active users, the maximum degree of anonymity is achieved when and attacker sees all the users equally probable as being the originator of a request (i.e., an anonymity set with N elements). In this model the degree of anonymity depends on the distribution of probabilities and not the number of users.

The model compares the information obtained by the attacker after observing the system against the optimal situation. It considers an attacker that can monitor all communication lines of the system, knows the number of active users and groups, the group a request originated from, and the number of requests produced by each user.

### 4.4. Reputation

Although not directly related with anonymous communication systems, Mislove et al. [21] explore the use of trust relationships such as social links, to thwart unwanted communications. It does so by bounding the total amount of unwanted communications that a user can produce, to the number of trust relationships that the user possesses. Receivers are responsible for reviewing and classify unwanted communications.

The proposed architecture relies on existing trust networks to connect senders and receivers via chains of pairwise trust relationships. It makes use of the fact that it is difficult for a user to create arbitrarily many trust relationships, thus it would be difficult for a malicious user to create enough trust relationships in order to flood the network with unwanted communications.

While not directly related to anonymous communication systems, Hogg et al. [22] also exploit the use of social networks for reputation systems. It has a relevant impact for anonymous communication systems since reputation systems are also highly affected by collusion among participants and Sybil attacks.

The authors present several benefits that social networks can have on reputation systems:

- Rating based on individuals' position in the social network, without the need for explicit user intervention.
- Ratings based on the social group of the individual being rated.
- Allow users to select among various filters gives flexibility in using social networks which can complicate attempts to distort the reported reputations.
- The use of social networks makes it more difficult to spoof the system by creating false identities or colluding in small groups.
- Large scale analysis of social networks can uncover certain forms of group collusion.

While the work presented so far assumes that there already exists a social network, and mainly focus on the use of trust for path selection and attack prevention, this work provides interesting results that could be used for the creation of the social network by a new user. Reputation could be used as a metric to rate nodes already in the social network to provide enough information for new nodes joining the network to bootstrap their social network.

## 5. Conclusions

We have shown several proposals that take advantage of social networking features to enhance the anonymity provided by anonymous communication systems, as well as to prevent several known attacks that are still possible in nowadays deployed systems.

While most of the work presented assumes that a social network already exists, and focuses on trust and how it can be used to build traffic analysis resistant communication paths, there are other interesting proposals that deal with reputation, which provides the means to rate the behavior of nodes in the network. This particularly important in bootstrapping a new social network, for instance, when a new user joins the systems and needs to find some trusted nodes to fit its needs.

Sybil attacks, which have been a long standing problem in de-centralized systems that provide no means to verify the identity of the entity running the nodes, can now be prevented through the use of social networks.

As in any other social network, it is expected that, if applied to anonymous communication systems, a profile would be created by each user of the social network. While it is true that this can have several implications in the users' anonymity, and needs further research, it was shown through a information theory context that the extra information provided by a user profile does not necessarily reduce the anonymity of the user.

As future work it would be interesting to map the requirements and design challenges of a social network on top of the available anonymous communication systems.

## References

[1] M. K. Reiter and A. D. Rubin, "Crowds: anonymity for web transactions," *ACM Trans. Inf. Syst. Secur.*, vol. 1, no. 1, pp. 66–92, 1998.

[2] G. Danezis and A. Serjantov, "Statistical disclosure or intersection attacks on anonymity systems," in *Proceedings of 6th Information Hiding Workshop (IH 2004)*, LNCS, (Toronto), May 2004.

[3] B. N. Levine, M. K. Reiter, C. Wang, and M. K. Wright, "Timing attacks in low-latency mix-based systems," in *Proceedings of Financial Cryptography (FC '04)* (A. Juels, ed.), pp. 251–265, Springer-Verlag, LNCS 3110, February 2004.

[4] J. R. Douceur, "The sybil attack," in *IPTPS '01: Revised Papers from the First International Workshop on Peer-to-Peer Systems*, (London, UK), pp. 251–260, Springer-Verlag, 2002.

[5] D. L. Chaum, "Untraceable electronic mail, return addresses, and digital pseudonyms," *Commun. ACM*, vol. 24, no. 2, pp. 84–90, 1981.

[6] D. M. Goldschlag, M. G. Reed, and P. F. Syverson, "Hiding routing information," in *Proceedings of the First International Workshop on Information Hiding*, (London, UK), pp. 137–150, Springer-Verlag, 1996.

[7] R. Dingledine, N. Mathewson, and P. Syverson, "Tor: The second-generation onion router," in *Proceedings of the 13th USENIX Security Symposium*, (San Diego, CA, USA), pp. 303–320, August 2004.

[8] I. Clarke, O. Sandberg, B. Wiley, and T. W. Hong, “Freenet: a distributed anonymous information storage and retrieval system,” in *International workshop on Designing privacy enhancing technologies*, (New York, NY, USA), pp. 46–66, Springer-Verlag New York, Inc., 2001.

[9] “Anonymous web surfing and online anonymity solutions — anonymizer,” *http://www.anonymizer.com/*. Last Checked July 2010.

[10] M. Wright, M. Adler, B. N. Levine, and C. Shields, “Defending anonymous communication against passive logging attacks,” in *Proceedings of the 2003 IEEE Symposium on Security and Privacy*, pp. 28–43, May 2003.

[11] M. Wright, M. Adler, B. N. Levine, and C. Shields, “The predecessor attack: An analysis of a threat to anonymous communications systems,” *ACM Transactions on Information and System Security (TISSEC)*, vol. 4, pp. 489–522, November 2004.

[12] P. Syverson, G. Tsudik, M. Reed, and C. Landwehr, “Towards an analysis of onion routing security,” in *Proceedings of Designing Privacy Enhancing Technologies: Workshop on Design Issues in Anonymity and Unobservability* (H. Federrath, ed.), pp. 96–114, Springer-Verlag, LNCS 2009, July 2000.

[13] K. Puttaswamy, A. Sala, and B. Y. Zhao, “Improving anonymity using social links,” in *Secure Network Protocols, 2008. NPSec 2008. 4th Workshop on*, pp. 15–20, 19-19 2008.

[14] G. Danezis, C. Diaz, C. Troncoso, and B. Laurie, “Drac: an architecture for anonymous low-volume communications,” in *Proceedings of the 10th international conference on Privacy enhancing technologies*, PETS’10, (Berlin, Heidelberg), pp. 202–219, Springer-Verlag, 2010.

[15] A. Johnson and P. Syverson, “More anonymous onion routing through trust,” in *Proceedings of the 2009 22nd IEEE Computer Security Foundations Symposium*, (Washington, DC, USA), pp. 3–12, IEEE Computer Society, 2009.

[16] H. Yu, M. Kaminsky, P. B. Gibbons, and A. Flaxman, “Sybilguard: defending against sybil attacks via social networks,” *SIGCOMM Comput. Commun. Rev.*, vol. 36, pp. 267–278, August 2006.

[17] H. Yu, P. B. Gibbons, M. Kaminsky, and F. Xiao, “Sybillimit: A near-optimal social network defense against sybil attacks,” in *Proceedings of the 2008 IEEE Symposium on Security and Privacy*, (Washington, DC, USA), pp. 3–17, IEEE Computer Society, 2008.

[18] C. Diaz, C. Troncoso, and A. Serjantov, “On the impact of social network profiling on anonymity,” in *Privacy Enhancing Technologies* (N. Borisov and I. Goldberg, eds.), vol. 5134, ch. Lecture Notes in Computer Science, pp. 44–62, Springer Berlin / Heidelberg, 2008.

[19] C. Diaz, C. Troncoso, and G. Danezis, “Does additional information always reduce anonymity?,” in *Proceedings of the Workshop on Privacy in the Electronic Society 2007* (T. Yu, ed.), (Alexandria,VA,USA), pp. 72–75, ACM, October 2007.

[20] S. S. Claudia Diaz, Joris Claessens and B. Preneel, “Information theory and anonymity,” in *Werkgemeenschap voor Informatie en Communicatietheorie* (B. Macq and J. J. Quisquater, eds.), pp. 179–186, 2002.

[21] A. Mislove, A. Post, P. Druschel, and K. P. Gummadi, “Ostra: leveraging trust to thwart unwanted communication,” in *Proceedings of the 5th USENIX Symposium on Networked Systems Design and Implementation*, NSDI’08, (Berkeley, CA, USA), pp. 15–30, USENIX Association, 2008.

[22] T. Hogg, “Security challenges for reputation mechanisms using online social networks,” in *Proceedings of the 2nd ACM workshop on Security and artificial intelligence*, AISec ’09, (New York, NY, USA), pp. 31–34, ACM, 2009.