Journal of Universal Computer Science, vol. 15, no. 9 (2009), 1797-1811
submitted: 26/7/08, accepted: 25/4/09, appeared: 1/5/09 © *J.UCS*

A Resilient P2P Anonymous Routing Approach Employing Collaboration Scheme

Junzhou Luo
(School of Computer Science and Engineering, Southeast University, Nanjing, P.R. China
jluo@seu.edu.cn)

Xiaogang Wang
(School of Computer Science and Engineering, Southeast University, Nanjing, P.R. China
Changzhou College of Information Technology, Changzhou, China
wxiaog@seu.edu.cn)

Ming Yang
(School of Computer Science and Engineering, Southeast University, Nanjing, P.R. China
yangming2002@seu.edu.cn)

Abstract: Node churn is a hindrance to construction of P2P-based anonymous networks, which makes anonymous paths fragile and results in message losses and communication failures. A collaboration scheme combining Friendly Neighbor-based Incentive (FNI) and Re-encryption mechanism is proposed to deal with the high node churn characteristic of P2P networks. The FNI mechanism encourages peers to forward other peers' messages, and establishes more connections to improve the performance of P2P networks, where only stable and well-behaved nodes can be chosen as relay nodes to improve the durability of anonymous paths. The Re-encryption mechanism is designed to replace those failed relay nodes and achieve routing resilience upon different node availabilities in real-world systems. The results from security analysis and simulation show that the P2P anonymous routing approach employing collaboration scheme significantly improves routing resilience and maintains low latency and modest communication overhead.

Keywords: anonymous routing, Peer-to-Peer, FNI mechanism, Re-encryption mechanism
Categories: C.2.0, C.2.4, C.2.6

1 Introduction

Many online applications such as e-banking, electronic voting, information sharing and searching etc, require anonymous measures to prevent third parties from gathering online private information. Most anonymous protocols proposed can be categorized into mix-based and multicast-based types. Mix-based protocols achieve anonymity by applying the redirection technique, where messages from the initiator are routed through a set of relay nodes (called mixes) until they reach their destination. Multicast-based protocols, in contrast, achieve anonymity by employing multicasting technique, where initiators and responders first join multicast groups and their messages are then multicasted to all group members.

Recently, there has been a growing interest in the construction of P2P-based anonymous networks. P2P systems can provide the open set of peer nodes offering a

potentially large anonymity set for participants, and can address the scalability issue that has been a challenge to static anonymous networks operating a small number of fixed mixes. Communication patterns and heterogeneity of peer nodes' locations also render the P2P network an appealing environment suitable for hiding anonymous traffic. Especially, mix-based protocols have been widely used in P2P anonymous systems.

Although P2P networks present a promising approach to constructing anonymous system, the node churn (changes in system membership) in P2P networks occurs often and becomes a hindrance to using the P2P network as an attractive environment for anonymization. A study by Saroiu et al. [Saroiu et al. 2002] has shown that P2P networks exhibit high node churn, where peers frequently leave or join the network and most peers are connected to overlay networks for a short period of time. The node churn complicates anonymous path construction in mix-based protocols, which makes anonymous paths fragile and short-lived. The failures of relay nodes on a routing path will also disrupt the path, resulting in message losses (including requests and responses) and awful user experiences. Especially, short-lived paths cannot support long-standing communication sessions. The membership churn is mainly caused by dynamic node lifetime which can be modelled as the Pareto distribution, and various forms of abuses and attacks such as free riders and denial-of-service (DoS) attacks, because the system may be lacking any "viable" incentive mechanism that encourages users to behave in the best interest of the community.

An intuitive solution to fragile paths in mix-based protocols is to apply broadcasting or multicasting technique. Although multicasting technique can mask node or link failures and improve path resilience, it incurs costly bandwidth consumption due to the massive messages and the cover traffic used to hide anonymous traffic. TAP [Zhu et al. 2004] and Cashmere [Zhuang et al. 2005] utilize a group of nodes as a mix to mask node failures in anonymous routing, requiring group members to share group keys. However, the anonymity may be degraded as the secrecy can be easily abused by group members. The message redundancy scheme [Zhu et al. 2007] was also proposed to mask node failures by applying erasure coding and path redundancy approaches. However, besides the complexity of implementation, the performance is influenced by additional communication overhead.

We herein present a resilient anonymous routing approach via wise choice of relay nodes to construct anonymous paths and wise choice of backup nodes to replace failed relay nodes on the anonymous path. We propose FNI mechanism as an incentive scheme for the peer to dynamically rate its neighbor nodes according to their behavior, where in exchange for better services, peers are encouraged to forward other peers' messages, establish more connections and behave better to improve the performance of P2P networks, thereby improving the durability of anonymous paths. We also propose Re-encryption mechanism based on proxy re-encryption scheme [Ateniese et al. 2006] to replace failed relay nodes on the anonymous path.

The security analysis shows that our scheme can achieve required initiator anonymity. We demonstrate the effectiveness of our collaboration approach by simulation. Our biased mix choice method based on FNI mechanism significantly improves routing resilience compared to random mix choice method, and is more efficient and easier to be implemented than the message redundancy approach.

The rest of the paper is organized as follows. Section 2 reviews the existing P2P anonymous routing schemes. Our approach is presented in Section 3. We present the anonymity analysis of our approach in Section 4. In section 5, we employ the simulation to demonstrate the effectiveness of our approach. We conclude the paper with future research directions in Section 6.

2 Related Work

Most anonymous routing protocols can be categorized into mix-based and multicast-based types. Mix-based systems [Chaum 1981] achieve anonymity by applying the redirection technique such as Tor [Dingledine et al. 2004], while multicast-based systems achieve anonymity by employing multicasting technique such as P5 [Sherwood et al. 2005], Hordes [Shields et al. 2000] and APFS [Scarlata et al. 2001].

As P2P networks are becoming appealing platforms for constructing anonymous systems, a number of P2P-based anonymous systems have been proposed. Tarzan [Freedman et al. 2002] and MorphMix [Rennhard et al. 2002] achieve anonymity by applying layered encryption and multi-hop routing approaches. In contrast, Crowds [Reiter et al. 1998] achieves anonymity by applying probabilistic random forwarding scheme. Although P2P-based anonymous systems aim to achieve more anonymity, the performance is degraded due to the node churn in P2P networks, which makes anonymous paths fragile and short-lived.

To improve the performance of P2P networks, several incentive mechanisms have been proposed. [Sun et al. 2004] proposed a simple Selfish Link-based InCentive (SLIC) approach for unstructured P2P file sharing systems where nodes in exchange for better services are encouraged to share more data, give more capacity to neighbor nodes' queries, and add new overlay links. To rate neighbor nodes efficiently, several incentive mechanisms have been proposed recently. [Barth et al. 2008] proposed a complete distributed solution to the transit price negotiation problem with incomplete information. [Wu et al. 2008] proposed the approach to rank retrieval systems in the condition of partial relevance judgments. Such incentive mechanisms need to be studied further for being employed in anonymous network.

To make anonymous paths resilient to node failures, TAP [Zhu et al. 2004] and Cashmere [Zhuang et al. 2005] decouple paths from fixed relay nodes and use a group of nodes in structured P2P networks [Rowstron et al. 2001] to mask single relay node failures by sharing group keys among group members. The main limitation of such systems is that group members must share some secrecy such as group keys. In order to avoid such limitation, the re-encryption mechanism combined with key exchange protocols can be used. [Ateniese et al. 2006] proposed an improved proxy re-encryption scheme allowing third-parties (proxies) to alter a ciphertext which has been encrypted for one party to be decrypted later by another. [Jeong et al. 2008] proposed an efficient parallel key exchange protocol among multiple parties and [Hwang et al. 2008] proposed the secure CL-PKE scheme to withstand malicious key generation centre attack.

MuON [Bansod et al. 2008] is a mutual anonymity system that uses epidemic-style data dissemination to handle network dynamics in unstructured P2P networks. Zhu et al. [Zhu et al. 2007] proposed a message redundancy scheme to mask node failures by employing erasure coding and path redundancy approaches. Although such

schemes achieve resilience upon node failures, the performance is influenced due to more additional communication overhead.

3 Anonymous Routing Approach

We propose an anonymous routing approach relying on Public Key Infrastructure (PKI), and assume that each node in P2P networks can learn its neighbor nodes' public keys through some mechanism (e.g., out-of-band or piggybacking). Our approach employs gossip protocol to manage nodes' states, by which each node learns information about its neighbor nodes and uses a subset of the known nodes to construct anonymous paths. The approach achieves initiator anonymity by applying layered encryption and multi-hop routing techniques through a set of relay nodes.

Two steps are followed to make anonymous routing resilient upon node failures. We first choose stable and well-behaved peers as relay nodes based on FNI mechanism. To get better service, peers are encouraged to forward other peers' messages and establish more connections to improve the performance of P2P networks, thereby improving the durability of the anonymous path. We then choose backup nodes to replace those failed relay nodes based on Re-encryption mechanism, so as to maximize routing resilience upon different node availabilities in real-world systems.

3.1 Design Goals

The collaboration scheme aims to make P2P anonymous routing resilient while maintaining low latency and modest communication overhead. A wide pool of neighbor nodes may be chosen to relay each other's traffic so as to gain anonymity. The main goals are described as follows.

Initiator anonymity: The identity of an initiator of communication is hidden to all nodes except the initiator itself.

Routing resilience: Anonymous routing is fault-tolerant against node or link failures, and messages can be delivered reliably.

Low latency: Anonymous messages can be delivered at low latency.

Low bandwidth cost: Routing resilience does not incur costly bandwidth consumption.

In this paper, we mainly focus on the routing resilience issue for initiator anonymity, since responder anonymity and mutual anonymity can be easily achieved by extending our design, i.e., using an additional level of redirection. We assume that the attacker aiming to break others' anonymity controls a fraction of nodes that can collude and share information with each other. The attacker can observe some fraction of network traffic and there is zero latency of messages between compromised nodes.

3.2 FNI Mechanism

FNI is an incentive mechanism suitable for P2P networks. When an initiator wants to construct anonymous paths, it will choose its neighbor node with high rank. Based on FNI mechanism, the rank of neighbor node is rated dynamically according to its behaviour. Each node can rate its neighbor nodes and use the rank to control how

many queries from each neighbor to be processed and forwarded on. When a neighbor provides more service than previously expected, the corresponding rank will increase. Similarly, less service will result in lower rank. In other words, FNI mechanism employs this mutual access control relationship as means of retaliation when a node does not play fair or connects to nodes that do not play fair. The nodes with higher rank are rewarded with more publicity and possibly better anonymity. To improve their services, nodes are encouraged to forward messages or to connect with nodes that have high capabilities. Only when a node's neighbors give it a high rank will it receive better service. There are four options for a node to increase its rank valuated by its neighbors and get better service.

1) Increasing answering power. By forwarding more queries or messages, a node can become more attractive.

2) Increasing the number of edges (or connectivity). By having more edges, a neighbor's queries can be forwarded to more nodes, which will lead to more hits for neighbors' queries.

3) Increasing the capacity of serving neighbors' queries. By donating more capacity, a node can forward more queries and improve the performance of the network.

4) Increasing the duration of serving neighbors' queries by increasing its durability.

Informally, FNI mechanism is a general mechanism that operates in periods, e.g., every minute. During each period, a node has certain capacity that it is willing to use for serving queries from neighbor nodes in P2P networks. To distinguish good neighbors from bad ones, a node u maintains a rank $R(u, v_i)$ $(1 \leq i \leq m)$ for each neighbor v_i, where $0 \leq R(u,v_i) \leq 1$. A rank of 1 indicates an excellent neighbor or close friend while a rank of 0 implies a useless one. With these ranks, a node then allocates its capacity to serve incoming queries from its neighbors proportionally to the rank. For instance, if node u has two links to nodes x and y with ranks 1 and 0.5 respectively, then in this period node u will give 2/3 of its capacity to queries from node x and 1/3 of its capacity to queries from node y. At the end of a period, each node reevaluates its opinion or ranks of its neighbors based on how much service the neighbors had provided during the current period. Since quality of service may fluctuate frequently, FNI mechanism employs an exponential decay mechanism to update ranks.

Compared to other concentrated reputation system in P2P networks, FNI mechanism allows each node to keep statistics about its neighbors and do not rely on a trusted authority or others to give accurate "reputations" about unknown nodes. Each node takes its neighbors with high rank as its friends. When node u wants to choose a relay node from its neighbors to construct a routing path, node u will choose its neighbor with higher rank as its successor to improve the stability of the path.

3.3 Anonymous Path Construction

Existing mix-based anonymous protocols do not take into account node behavior when choosing mixes to construct anonymous paths. If nodes on a path are prone to fail or leave, the resulting path is fragile and short-lived. Intuitively, the durability and quality of the anonymous path can be improved by choosing well-behaved nodes as

mixes. So we propose a biased mix choice algorithm to wisely choose relay nodes for the purpose of prolonging path lifetime.

The anonymous path or tunnel is constructed incrementally. The initiator is responsible for constructing onion-encrypted connections relayed through a sequence of intermediate nodes. All participating nodes run software that 1) discovers its participating neighbors, 2) intercepts packets generated by local applications that should be anonymized, 3) manages tunnels through chains of other nodes to anonymize these packets, 4) forwards packets to implement other nodes' tunnels.

Typical anonymous communication proceeds in following stages.

1) **Relay nodes choosing.** An initiator u running an anonymous application first chooses its successor v from its neighbors based on FNI mechanism. Then the initiator u ask node v to recommend v's successor w from v's neighbors, and then the initiator u ask node w to recommend w's successor s from w's neighbors, and this process goes on until the initiator u finishes choosing required routing relay nodes through the overlay network.

2) **Anonymous path or tunnel establishing.** The initiator u constructs an anonymous path by applying layered encryption approach. Each node on the path removes or adds a layer of encryption, depending upon the packet's direction of traversal. Let $P_i(1 \leq i \leq L)$ be a relay node of a forwarding path with length L and P_{L+1} be the responder D. This source-routing node u uses these chosen relay nodes to establish a tunnel, which includes the distribution of session keys $R_i(1 \leq i \leq L)$. The session key R_i is used for symmetric encryption and decryption between the initiator u and the relay node P_i. The initiator u first generates a forwarding path onion containing session keys for each relay node as follows.

$$Path_i = \begin{cases} \perp\ (termination) & i = L+1 \\ < P_{i+1}, R_i, Path_{i+1} > EnPubKey_{P_i} & 1 \leq i \leq L \end{cases}$$

Then the initiator u generates a random stream ID sid_0, caches the tuple $[sid_0, P_1]$, and sends the tuple $[Path_1, sid_0]$ to the first relay node P_1. In general, upon receiving the message $[Path_i, sid_{i-1}]$ from the upstream node, the i-th relay node P_i on the anonymous path strips off the outer layer of $Path_i$ using its private key $PriKey_{P_i}$, revealing the next hop P_{i+1} and symmetric key R_i. If $Path_{i+1}$ is not $\perp$ then the i-th relay node generates a random stream sid_i, caches the tuple $[P_{i-1}, sid_{i-1}, P_{i+1}, sid_i, R_i]$ and sends $[Path_{i+1}, sid_i]$ to P_{i+1}. Otherwise, it caches the tuple $[P_{L-1}, sid_{L-1}, D, \perp, R_L]$ implying the end of the forwarding path.

3) **Path states backuping:** In order to be resilient to node failure, node P_i on the anonymous path caches the tuple $[P_{i-1}, sid_{i-1}, P_{i+1}, sid_i, R_i]$ (denoted as $CachR_i$) while encrypting the cached information $CachR_i$ into $Backup_i$ with its public key $PubKey_{P_i}$.

$$Backup_i = < CachR_i > EnPubKey_{P_i} = < [P_{i-1}, sid_{i-1}, P_{i+1}, sid_i, R_i] > EnPubKey_{P_i}$$

Each mix node along the anonymous path generates corresponding re-encryption keys with its friendly neighbors. Let $N_{ij}(1 \leq j \leq m)$ be friendly neighbors of the relay node P_i. By applying our Re-encryption mechanism, the relay node P_i can first

generate its delegation (re-encryption) key $ReenKey_{ij}$ ($1 \le j \le m$) with its friendly neighbor N_{ij}. The relay node then delivers the tuple [$Backup_i$, N_{ij}, $ReenKey_{ij}$ ($1 \le j \le m$)] containing its corresponding re-encryption keys and cached information to its predecessor P_{i-1}. The tuple [$Backup_i$, N_{ij}, $ReenKey_{ij}$ ($1 \le j \le m$)] will be used for failure recovery later.

4) **Data delivering anonymously.** Finally, the initiator u routes packets through this tunnel. The exit point of the tunnel is the tailed node P_L, which forwards the anonymized packets to servers that are not aware of these chosen mixes. P_L also receives the response packets from these servers and reroutes the packets back over this tunnel. The initiator u generates the payload as follows for each tunnel while delivering packets anonymously.

$$PayLoad_i = \begin{cases} message & i = L+1 \\ < PayLoad_{i+1} > EnR_i & 1 \le i \le L \end{cases}$$

The initiator u first looks for cached sid_0 from its cached tuple [sid_0, P_1] and then sends the tuple [sid_0, $PayLoad_1$] to the first relay node P_1. In general, upon receiving [sid_{i-1}, $PayLoad_i$], node P_i first looks for the corresponding symmetric key R_i, stream ID sid_i and the next node P_{i+1} according to the received sid_{i-1}. Then it strips off the outer layer of $PayLoad_i$ by using R_i. Unless P_{i+1} is D, node P_i will forward $PayLoad_{i+1}$ to node P_{i+1}. Finally, node P_L sends the message to the responder D. Once the responder D sends the response message back to the initiator, the payload is encrypted by the cached symmetric key at each relay node until reaching the initiator who strips the onion and gets the response message.

3.4 Re-encryption Mechanism

In 1998, Blaze, Bleumer, and Strauss (BBS) proposed a cryptographic application called *atomic proxy re-encryption,* in which a semi-trusted proxy converts a ciphertext for Alice into a ciphertext for Bob without seeing the underlying plaintext. [Ateniese et al. 2006] proposed an improved proxy re-encryption scheme which has the following properties.

1) **Unidirectional.** Delegation from A→ B does not allow re-encryption from B→ A.

2) **Non-interactive.** Re-encryption keys can be generated by Alice using Bob's public key, no trusted third party or interaction is required.

3) **Proxy invisibility.** Both the sender and recipient are aware of the proxy re-encryption protocol but do not know whether the proxy is active, or the proxy has performed any action or made any changes, or even if it exists (the proxy is indeed "invisible"). More specifically, we allow the sender to generate an encryption that can be opened only by the intended recipient (first-level encryption) or by any of the recipient's delegatees (second-level encryption). At the same time, we can ensure that any delegatee will not be able to distinguish a first-level encryption (computed under his public key) from a re-encryption of a ciphertext intended for another party (we are assuming that the encrypted message does not reveal any information that would help the delegatee to make this distinction).

4) **Non-transitive.** The proxy alone cannot re-delegate decryption rights.

The re-encryption mechanism can be implemented over two groups G_1, G_2 of prime order q with a Tate pairing bilinear map $e: G_1 \times G_1 \to G_2$. The system parameters are random generators $g \in G_1$ and $Z=e(g,g) \in G_2$. There are several basic building blocks in the re-encryption mechanism.

1) **Key pair generating.** Alice's key pair can be generated in the form $pk_a=g^a$, $sk_a=a$.

2) **Re-encryption key generating.** Alice can delegate its decryption right to Bob by publishing the re-encryption key $\mathrm{rk}_{\mathrm{A}\to\mathrm{B}}=\mathrm{g}^{\mathrm{b/a}} \in \mathrm{G}_1$ which can be computed from B's public key $pk_b=g^b$.

3) **First-level encrypting.** To encrypt a message $m \in G_2$ under pk_a in such a way that it can only be decrypted by the holder of sk_a, output $c=(Z^{ak}, mZ^k)$.

4) **Second-level encrypting.** To encrypt a message $m \in G_2$ under pk_a in such a way that it can only be decrypted by Alice and her delegates, output $c=(g^{ak}, mZ^k)$.

5) **Re-encrypting.** The third party can change a second-level ciphertext for Alice into a first-level ciphertext for B with $\mathrm{rk}_{\mathrm{A}\to\mathrm{B}}=\mathrm{g}^{\mathrm{b/a}}$. Having $c_a=(g^{ak}, mZ^k)$, compute $e(g^{ak}, g^{b/a})=Z^{bk}$ and publish $c_b=(Z^{bk}, mZ^k)$.

6) **Decrypting.** To decrypt a first-level ciphertext $c_a=(\alpha,\beta)$ with secret key $sk=a$, we can compute $m=\beta/\alpha^{1/a}$. To decrypt a second-level ciphertext $c_a=(\alpha,\beta)$ with secret key $sk=a$, we can compute $m=\beta/e(\alpha,g)^{1/a}$.

In our Re-encryption mechanism, the relay node P_i first generates its delegation (re-encryption) key $ReenKey_{ij}$ ($1 \le j \le m$) with its friendly neighbor N_{ij}.

$$ReenKey_{ij} = rk_{P_i \to N_{ij}} = g^{sk_{N_{ij}} / sk_{P_i}} \in G_1 \quad (1 \le j \le m)$$

The relay node P_i then encrypts its cached information $CachR_i$ in second-level encryption form by using its public key $PubKey_{P_i}$.

$$Backup_i = < CachR_i > EnPubKey_{P_i}$$

Finally the relay node P_i delivers the tuple [$Backup_i$, N_{ij}, $ReenKey_{ij}$ ($1 \le j \le m$)] containing its corresponding re-encryption keys and second-level encrypted cached information to its predecessor P_{i-1}. The tuple [$Backup_i$, N_{ij}, $ReenKey_{ij}$ ($1 \le j \le m$)] will be used for failure recovery later.

3.5 Failure Node Detection and Substitution

A path consisting of a sequence of relay nodes is disrupted if any node on the path fails. Each node seeking anonymity needs to maintain a cache that keeps track of its neighbor nodes so that it can pick cached neighbors as mixes to construct anonymous paths. Based on FNI mechanism, each node can maintain its neighbors' current states. When node P_i finds its neighbors' membership changed by piggybacking node liveliness information onto P2P protocol messages, the relay node P_i will update its cached neighbors' states accordingly.

When the relay node P_{i-1} finds its successor's membership changed on the anonymous path (e.g., node P_i failed or left), node P_{i-1} will choose a candidate neighbor node N_{is} with high rank from its cached neighbors, as node N_{is} is also a friendly neighbor of node P_{i-1} by looking up its stored path states delivered by its successor P_i, then node P_{i-1} will re-encrypt $Backup_{\mathrm{i}}$ originally encrypted by P_i's public key to $Backup_{\mathrm{is}}$ encrypted by N_{is}'s public key with re-encryption key $ReenKey_{\mathrm{is}}$. Meanwhile, node P_{i-1} caches its modified tuple $[P_{i-2}, sid_{i-2}, N_{is}, sid_{i-1}, R_{i-1}]$. Node N_{is} will be able to decrypt the $Backup_{\mathrm{is}}$ with its own private key, and get the correct cached anonymous routing information originally stored in node P_i. Finally, node N_{is} notifies its successor P_{i+1} to modify cached tuple as $[N_{is}, sid_i, P_{i+2}, sid_{i+1}, R_{i+1}]$ according to sid_i. The anonymous path remains uninterrupted as the failed node P_i is successfully substituted. Without changing the initiator's routing path, the initiator can still use original constructed routing path from its view. So, the resilience of routing holds.

We address the routing resilience issue of anonymity protocols from two sides. On the one hand, we propose FNI as an incentive scheme to encourage peers to behave well and provide more stable service. We base our mix choices on neighbors' ranks and pick nodes that tend to stay longer and behave well as mixes, resulting in more resilient routing paths. On the other hand, we can mask node failures to improve the quality of P2P networks by using Re-encryption mechanism while minimizing bandwidth cost.

4 Security Analysis

Our approach achieves initiator anonymity essentially by utilizing onion routing mechanism [Dingledine et al. 2004]. Even in possession of neighbors' rank information, the attacker (except the responder) is difficult to break the anonymity by observing some traffic on the anonymous path. Our collaboration scheme ensures that only stable and well-behaved neighbors can be chosen to substitute failed relay nodes. Communication patterns and heterogeneity of peer nodes' locations in P2P networks, together with cover traffic, complicates the statistical attack. Message confidentiality is achieved by symmetric encryption.

To analyze the anonymity of our scheme, we assume that the attacker occupies one or more positions on a path initiated by a non-malicious node and all other non-malicious nodes can act as the initiator with equal probability. The goal of the attacker is to determine the node who initiates the path. Since messages are encrypted, the attacker has no reason to suspect any node other than the one immediately preceding it. We define initiator anonymity as the probability that the immediate predecessor of the first malicious node on the path is in fact the initiator. We analyze the initiator anonymity with respect to four parameters: n (total number of candidate nodes in the system), m (average number of neighbor nodes of each node), p (fraction of non-malicious nodes over all candidate nodes), and L (number of relay nodes on a path). We employ the approach proposed by Wang [Wang 2004] to analyze the initiator anonymity of our scheme. Therefore, the anonymity degree of any initiator can be denoted as $\Pr[I|H_{1+}]$ representing the conditional probability that the predecessor of

the first malicious node is indeed the initiator, given that the first malicious node is occupying the first position or somewhere else after the first position.

We first present some notations in order to facilitate the analysis of our scheme. Let *I* denote the event that the predecessor of the first malicious node on the path is indeed the initiator. H_i *(1 ≤ i ≤ L)* refers to the event that the first malicious node on the path is occupying the *i*-th position on the path, while H_{i+} *(1 ≤ i ≤ L)* refers to the event that the first malicious node may occupy the *i*-th position or somewhere else after the *i*-th position on the path.

Since the probability that the first malicious node on the path occupies the *i*-th position is calculated as follows: $\Pr[H_i] = p^{i-1}(1-p)$, then we can calculate the probability that the first malicious node occupies the first position or somewhere else after the first position as follows:

$$\Pr[H_{1+}] = \sum_{i=1}^{L} \Pr[H_i] = \sum_{i=1}^{L} p^{i-1}(1-p) = 1 - p^L .$$

Given that the first malicious node occupies the *i*-th position on the path, the conditional probability that the predecessor of the first malicious node on the path is indeed the initiator can be calculated as follows:

$$\Pr[I \mid H_i] = \frac{1}{np} \cdot \frac{1 - (\frac{1}{mp} - \frac{1}{np})^{i-1}}{1 - (\frac{1}{mp} - \frac{1}{np})} + (\frac{1}{mp} - \frac{1}{np})^{i-1} \qquad (1 \le i \le L).$$

Therefore, we can calculate the probability that the predecessor of the first malicious node on the path is indeed the initiator *I* as follows:

$$\Pr[I] = \sum_{i=1}^{L} \Pr[H_i]\Pr[I \mid H_i]$$

$$= (1-p) + (1-p)\sum_{i=2}^{L} p^{i-1}(\frac{1}{np} \cdot \frac{1 - (\frac{1}{mp} - \frac{1}{np})^{i-1}}{1 - (\frac{1}{mp} - \frac{1}{np})} + (\frac{1}{mp} - \frac{1}{np})^{i-1}) .$$

Provided that there is at least one malicious node on the anonymous path, the anonymity of our scheme can be calculated as follows:

$$\Pr[I \mid H_{1+}] = \frac{\Pr[I \wedge H_{1+}]}{\Pr[H_{1+}]} = \frac{\Pr[I]}{\Pr[H_{1+}]}$$

$$= \frac{\frac{1}{n}}{p + \frac{1}{n} - \frac{1}{m}} + \frac{p - \frac{1}{m}}{(p + \frac{1}{n} - \frac{1}{m})\sum_{i=0}^{L-1} p^i}(1 + \sum_{i=1}^{L-1}(\frac{1}{m} - \frac{1}{n})^i) \qquad (1)$$

According to (1), the anonymity of our scheme increases as the average number of neighbor nodes of each node *m* increases. When *m* gets close to *n*, the conditional probability $\Pr[I|H_{I+}]$ representing the anonymity of random mix choices scheme can be approximated as follows:

$$\Pr[I \mid H_{1+}] = \frac{\Pr[I]}{\Pr[H_{1+}]} = \frac{(1-p) + \frac{1-p^{L-1}}{n}}{1-p^{L}} \quad (2)$$

Although the anonymity of our scheme can be increased by choosing larger value of *m*, larger value of *m* will degrade the efficiency of our scheme. The parameters of our scheme should be configured to balance the tradeoffs among the anonymity and efficiency.

5 Simulation

5.1 Simulation Setup

Simulator. We evaluate the effectiveness of our scheme by applying the P2P network simulator P2Psim. P2Psim supports *OneHop* [Gupta et al. 2004] routing protocol designed for structured P2P networks. Since *OneHop* protocol uses a hierarchical gossip protocol containing slice leaders, unit leaders and unit members to disseminate membership changes quickly and efficiently among nodes, each node can maintain a full routing table containing liveness information about every other node in the overlay. We augment *OneHop* protocol by piggybacking neighbors' rank information onto the gossip messages for biased mix choice. The initiator can choose the relay node by random mix choice or biased mix choice based on our FNI mechanism to construct its anonymous routing path. The number of relay nodes on a path is L=3 by default, unless otherwise specified.

Simulation network. We choose the *E2EGraph* as the network topology. The simulated network consists of 1024 nodes with inter-node latencies derived from measuring the pairwise latencies of 1024 DNS servers on the Internet using the *King* method [Gummadi et al. 2002]. The average round-trip time for the simulation network is 198ms. Unless otherwise specified, our simulation results presented in the paper are based on this simulation network. To simulate the network churn, each node alternately leaves and rejoins the network following the Pareto distribution with median time of 1 hour (i.e., α=1 and β=1800 seconds). The interval between successive events for each node such as leaving and rejoining the network is called node's lifetime or session time.

FNI mechanism. We use dynamic rank for each node in the simulation. When a node joins the network, a dynamic rank will be assigned to the node according to its configured lifetime time. Each node will pick its neighbor node with higher rank and lower latency to construct its anonymous routing path. A node with high rank means that it will stay longer and behave better to relay other node's communication messages.

Re-encryption mechanism. We use JHU-MIT Proxy Re-cryptography Library (PRL) [Green et al. 2007] based on MIRACL cryptographic library to implement our Re-encryption mechanism. The PRL implements two proxy re-encryption schemes in C++ language. By making use of the Tate pairing in supersingular elliptic-curve groups, JHU-MIT Proxy Re-cryptography Library provides various required functions for re-encryption mechanism such as delegation (re-encryption) key generation and re-encryption functions.

Evaluation metrics. We employ following metrics to evaluate the performance of our scheme.

1) **Latency.** Measuring successful routing latency. A successful routing means that the responder can receive a copy of the original message successfully.

2) **Path setup success rate.** Measuring the success rate of anonymous routing path construction.

3) **Path durability.** Measuring path lifetime and implying routing resilience. For non-substituting approach, i.e., without our Re-encryption mechanism, path lifetime is terminated whenever any node on the path fails. For Re-encryption mechanism, path lifetime is terminated if no appropriate neighbor node can be used to substitute the failed node.

4) **Bandwidth.** Measuring the average bandwidth cost of anonymous routing approach by which the responder successfully receives a message.

5.2 Simulation Results

Path construction. In this set of experiments, we simulated node churn by employing the Pareto distribution with median time of 1 hour as described in Section 5.1. The simulation time in each experiment was 2 hours. After all the 1024 nodes joined the network during the first simulation hour, path construction events were generated by employing the exponential distribution with average inter-arrival time of 116 seconds. The number of total path construction events for each random and biased FNI mechanism respectively was about 16,000. Average latency and path setup success rates can be shown in [Tab. 1]. The first row refers to the results of random mix choice while the second represents the results of biased mix choice based on FNI mechanism.

Mix choice	Path setup success rates (%)	Average latency (milliseconds)
Random	3.1	408
FNI	98.6	106

Table 1: Average latency and path setup success rates with random and FNI mechanism respectively.

Two main observations can be made from [Tab. 1]: 1) The path setup success rate is significantly increased by applying biased mix choice based on FNI mechanism. Since the living node with high rank will be chosen as relay node to construct the anonymous routing path, the probability that all the three chosen neighbor nodes are alive is bigger than that of random mechanism, significantly reducing the number of attempts in path construction. 2) The average latency can be significantly decreased by applying biased mix choice based on FNI mechanism. Biased mix choice allows the initiator to construct lower latency anonymous path, thereby decreasing the path latency.

Path durability and communication overhead. In this set of experiments, we simulated node churn by employing the Pareto distribution with median time of 1 hour also. There were two nodes kept staying alive in the system throughout the simulation, where one node acts as an initiator and the other acts as a responder. After the first hour of simulation, path construction events were generated. The path

durability is evaluated by the duration of the continuing successful communication. The simulation time in each experiment was 2 hours. The durability of the constructed path and the average bandwidth of each node can be shown in [Tab. 2]. The first row refers to the results of random mix choice while the second represents the results of biased mix choice without substituting failed nodes, and the third row refers to the results of biased mix choice based on Re-encryption mechanism.

Failed nodes substituting	**Path durability (seconds)**	**Bandwidth (KB)**
Random	779	4.5
Non-substituting biased	1410	4.8
Re-encryption mechanism based	1950	5.7

Table 2: Path durability with random, non-substituting biased and Re-encryption mechanism based respectively.

Two main observations can be made from the results: 1) Biased mix choice increases path stability by picking long-lived nodes as relay nodes, thereby increasing the path durability. 2) The Re-encryption mechanism can effectively enhance path stability by substituting failed nodes at the cost of modest increased bandwidth cost. Since each node is required to encrypt its cached information and deliver the encrypted cached information along with its re-encryption keys to its predecessor, additional bandwidth increase is incurred for each node.

When compared to the message redundancy scheme [Zhu et al. 2007], our scheme exhibits the similar performance. Especially, the path setup success rate of our scheme can be increased from 3.1% to 98.6% based on random mix choice and FNI respectively, while the path setup success rate of the message redundancy scheme is increased from 4.98% to 96.24% based on random choice and biased choice respectively. Although our scheme exhibits similar performance to the message redundancy scheme, our scheme is easy to be implemented.

6 Conclusions and Future Work

In this paper, a collaboration scheme combining FNI and Re-encryption mechanism is proposed to make P2P anonymous routing resilient. The FNI mechanism aims to create an incentive scheme to improve the performance of P2P networks, thereby improving the durability of anonymous path. The Re-encryption mechanism aims to reduce cost and adjust routes to mask the failed node. Since rebuilding the entire path would be expensive, the working relay nodes are retained and only the failed relay nodes are replaced. Based on FNI and Re-encryption mechanism, the durability of anonymous paths in P2P anonymous networks can be significantly enhanced by masking mix failures. Biased mix choice provides an incentive for nodes seeking better service to stay longer in the system, which as a result improves the stability and anonymity of the system. The results from our security analysis and simulation show that the collaboration scheme can significantly improve routing resilience while maintaining predictable latencies, high anonymity, and low communication overhead.

Although our novel anonymous routing approach improves routing resilience, the anonymity guarantees may be degraded by malicious attackers trying to break our schemes. In biased mix choice based on FNI mechanism, nodes that have good rank and have been alive for longer time are more likely to be chosen as relay nodes. Attackers may also attempt to stay longer in the system with the hope of being chosen as relay nodes of many paths, and then collude with other malicious nodes on these anonymous paths to break the participants' anonymity. In the near future, we plan to study the cooperative scheme to fight against such colluded malicious nodes and manage to make the anonymous routing more resilient.

Acknowledgements

This work is supported by National Natural Science Foundation of China under Grant No. 60773103, China Specialized Research Fund for the Doctoral Program of Higher Education under Grant No. 200802860031, Jiangsu Provincial Natural Science Foundation of China under Grant No. BK2007708 and BK2008030, Jiangsu Provincial Key Laboratory of Network and Information Security under Grant No. BM2003201, Key Laboratory of Computer Network and Information Integration of Ministry of Education of China under Grant No. 93K-9 and International Science and Technology Cooperation Program of China.

References

[Ateniese et al. 2006] Ateniese, G., Fu, K., Green, M., Hohenberger, S.: Improved Proxy Re-Encryption Schemes with Applications to Secure Distributed Storage, ACM Transactions on Information and System Security, 9(1), 2006, 1-30.

[Bansod et al. 2008] Bansod, N., Malgi, A., Choi, B. K., Mayo, J.: MuON: Epidemic based mutual anonymity in unstructured P2P networks, The International Journal of Computer and Telecommunications Networking, 52(5), 2008, 915-934.

[Barth et al. 2008] Barth, D., Echabbi, L., Hamlaoui, C.: Optimal Transit Price Negotiation: The Distributed Learning Perspective, Journal of Universal Computer Science, 14(5), 2008, 745-765.

[Chaum 1981] Chaum, D. L.: Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms, Communications of the ACM, 24(2), 1981, 84–90.

[Dingledine et al. 2004] Dingledine, R., Mathewson, N., Syverson, P.: Tor: The Second-Generation Onion Router, In Proceedings of the 13th Conference on USENIX Security Symposium, San Diego, CA, USA, 9–13 August 2004, 21-38.

[Freedman et al. 2002] Freedman M. J., Morris, R.: Tarzan: A Peer-to-Peer Anonymizing Network Layer, In Proceedings of the 9th ACM Conference on Computer and Communications Security (CCS), Washington, USA, 18-22 November 2002, 193-206.

[Green et al. 2007] Green, M., Ateniese, G.: Identity-Based Proxy Re-encryption, In Proceedings of the 5th International Conference on Applied Cryptography and Network Security (ACNS'07), Zhuhai, China, 5-8 June 2007, 288-306.

[Gummadi et al. 2002] Gummadi, K. P., Saroiu, S., Gribble, S. D.: King: Estimating Latency between Arbitrary Internet End Hosts, In Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment (IMW), Marseille, France, 6-8 November 2002, 5-18.

[Gupta et al. 2004] Gupta, A., Liskov, B., Rodrigues, R.: Efficient Routing for Peer-to-Peer Overlays, In Proceedings of the 1th Symposium on Networked Systems Design and Implementation (NSDI), San Francisco, USA, 29-31 March 2004, 9-23.

[Hwang et al. 2008] Hwang, Y. H., Liu, J. K., Chow S. S.M.: Certificateless Public Key Encryption Secure against Malicious KGC Attacks in the Standard Model, Journal of Universal Computer Science, 14(3), 2008, 463-480.

[Jeong et al. 2008] Jeong, I. R., Lee, D. H.: Parallel Key Exchange, Journal of Universal Computer Science, 14(3), 2008, 377-396.

[Reiter et al. 1998] Reiter, M. K., Rubin, A. D.: Crowds: Anonymity for Web Transactions, ACM Transactions on Information and System Security, 1(1), 1998, 66–92.

[Rennhard et al. 2002] Rennhard, M., Plattner, B.: Introducing Morphmix: Peer-to-peer based Anonymous Internet Usage with Collusion Detection, In Proceedings of the Workshop on Privacy in the Electronic Society, Washington, USA, 18-22 November 2002, 91-102.

[Rowstron et al. 2001] Rowstron, A., Druschel, P.: Pastry: Scalable, Decentralized Object Location, and Routing for Large-Scale Peer-to-Peer Systems, In Proceedings of the 18th IFIP/ACM International Conference on Distributed System Platforms (Middleware), Heidelberg, Germany, 12-16 November 2001, 329–350.

[Saroiu et al. 2002] Saroiu, S., Gummadi, P. K., Gribble, S. D.: A Measurement Study of Peer-to-Peer File Sharing Systems, In Proceedings of Multimedia Computing and Networking, San Jose, CA, USA, 23-24 January 2002, 156-170.

[Scarlata et al. 2001] Scarlata, V., Levine, B. N., Shields, C.: Responder Anonymity and Anonymous Peer-to-Peer File Sharing, In Proceedings of IEEE International Conference on Network Protocols (ICNP 2001), Riverside, CA, 11-14 November 2001, 272-280.

[Shields et al. 2000] Shields, C., Levine, B. N.: A Protocol for Anonymous Communication over the Internet, In ACM Conference on Computer and Communications Security, Athens, 1-4 November 2000, 33-42.

[Sherwood et al. 2005] Sherwood, R., Bhattacharjee, B., Srinivasan, A.: P5: A protocol for Scalable Anonymous Communication, Journal of Computer Security, 13(6), 2005, 839-876.

[Sun et al. 2004] Sun, Q., Garcia-Molina, H.: SLIC: A Selfish Link-based Incentive Mechanism for Unstructured Peer-to-Peer Networks, In Proceedings of the 24th International Conference on Computing Systems(ICDCS), Tokyo, Japan, 24-26 March 2004, 506-515.

[Wang 2004] Wang, W. P.: Study on Performance and Scalability of Anonymous Communication System, PhD Thesis, Central South University, 2004.

[Wu et al. 2008] Wu, S., Crestani, F.: Ranking Retrieval Systems with Partial Relevance Judgements, Journal of Universal Computer Science, 14(7), 2008, 1020-1030.

[Zhu et al. 2004] Zhu, Y., Hu, Y.: TAP: A Novel Tunnelling Approach for Anonymity in Structured P2P Systems, In Proceedings of International Conference on Parallel Processing (ICPP), Montreal, Quebec, Canada, 15-18 August 2004, 21–28.

[Zhuang et al. 2005] Zhuang, L., Zhou, F., Zhao, B. Y., Rowstron, A.: Cashmere: Resilient Anonymous Routing, In Proceedings of the 2nd Symposium on Networked Systems Design and Implementation (NSDI), Boston, MA, USA, 2-4 May 2005, 301-314.

[Zhu et al. 2007] Zhu, Y., Hu, Y.: Making Peer-to-Peer Anonymous Routing Resilient to Failures, In Proceedings of the 21st International Parallel and Distributed Processing Symposium (IPDPS), Long Beach, CA, USA, 26-30 March 2007, 4228-4238.