TITLE:
A Traceable and Pseudonymous P2P Information Distribution System

AUTHOR(S):
Tsujio, Naoki; Okabe, Yasuo

CITATION:

ISSUE DATE:
2015

URL:
http://hdl.handle.net/2433/198788

RIGHT:
© 2015 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. この論文は出版社版ではありません。引用の際には出版社版をご確認ご利用ください。This is not the published version. Please cite only the published version.
A Traceable and Pseudonymous P2P Information Distribution System

Naoki Tsujio  
Graduate School of Informatics  
Kyoto University  
Email: tsujio@net.ist.i.kyoto-u.ac.jp

Yasuo Okabe  
Academic Center for Computing and Media Studies  
Kyoto University  
Email: okabe@media.kyoto-u.ac.jp

Abstract—Anyone can publish various kinds of information on the Internet almost freely, but in some cases such information distribution is inhibited by the authorities. In order to resist such censorship, there have been developed many anonymous information distribution systems such as Freenet and Tor, but some people argue that such a system may also be a hotbed of crime since scrupulous anonymity disturbs investigations. In this paper, an article distribution system is proposed, which protects pseudonymity of users from surveillance by the authorities as with existing anticensorship systems. As a novel point, the proposed system allows a user to trace the publisher of an article by cooperation of the users who have relayed the article. This will suppress criminal acts abusing pseudonymity in the system. On the other hand, it is difficult to trace the publisher for a single government or an organization alone unless it obtains cooperation of multiple users. The proposed system will therefore be able to avoid authoritarian censorship or surveillance by the authorities. The system adopts P2P architecture, and a user can publish articles to other users like Netnews. A published article is relayed by node to node and spreads over the network of the system. In order to trace the publisher of an article, a user records a relaying log when he relays an article. A relaying log contains information about the predecessor from whom the user received the article. A user can trace the publisher by gathering relaying logs. Each user has responsibility to determine whether to disclose a relaying log or not, considering the content of the article. If all users along the path from the publisher agree to cooperate in gathering their logs, they will be able to trace to the publisher. The performance of the system is discussed in evaluation, how users’ actions affect traceability, and what should users do if governments intervene in the system.

Keywords—peer-to-peer; censorship; privacy; traceable; information distribution;

I. Introduction

The Internet has become a worldwide communication infrastructure where you can publish various kinds of information to an unspecified large number of people and can also access such published information at any time. However, there are some cases in which such activities are inhibited by the authorities. For example, a government may block access to unfavorable information or may punish people who published criticisms of the government. Such actions are called Internet censorship and are continuously reported as “Enemies of the Internet” [1] by the Reporters Without Borders [2]. In 2013, Edward Joseph Snowden, a former staff of the National Security Agency (NSA) of the USA, leaked information about the PRISM program of the NSA, which conducts surveillance of Internet communication and collection of personal information [3]. It is widely discussed about such collecting personal information and invasion of privacy by governments.

There are many researches on protecting the freedom of expression and the privacy of users from Internet censorship or surveillance. Freenet [4] and Tor [5] are typical representatives. Using these technologies, one can distribute information with high anonymity, avoiding Internet censorship.

However, can it be said that Internet censorship or surveillance is always evil and should be eliminated? There is a difficult problem in it (Figure 1). If a criminal declaration is published on the Internet, most people will desire to trace the publisher and to prevent it. On the other hand, if a criticism of a government is published, someone who goes along with it may wish it would widely spread. Considering these examples, we suppose that Internet censorship or surveillance are regarded as evil when they are conducted nevertheless many people dissent from it.

In this paper, we propose an information distribution system which is not affected by authoritarian censorship or surveillance by a single subject such as a government or a company. The system is designed as a standard article distribution system like Netnews [6] and allows users to publish articles pseudonymously. The proposed system allows a user to trace the distribution path of an article in specific cases. It is a clue for detecting the publisher of the article. The trace needs cooperation of all users who have relayed the article. The system constructs a Peer-to-Peer (P2P) network and spreads published articles by relaying between peers. When a peer relays an article, it records a relaying log in its local storage so that a user can trace the distribution path later. If all users along the path from the publisher agree to cooperate in gathering their logs, they will be able to trace to the publisher.

The rest of this paper is organized as follows. Chapter II describes related work and the difference from it. Chapter III shows the design of the system. Chapter IV discusses the evaluation of the system. Chapter V concludes this paper and mentions future work.

II. Related Work

A. P2P Information Distribution

There are some researches on defending the privacy of users and the freedom of expression. They allow users to publish information anonymously, and they are designed as
a distributed system, which has no Single Point of Failure (SPOF). In addition, the case of the NSA indicates that if a system is operated by a specific organization, a government can conduct surveillance by intervening in it. These systems are not operated by a specific administrator.

Freenet [4] is software by which a user can share files, browse, and publish Web pages anonymously. It is developed to defend the freedom of expression on the Internet. Devices running Freenet construct an overlay network and share their machine resource to provide functions of Freenet. Communication in Freenet is anonymized with encryption and multiple proxies, so that nobody can find who accesses what content.

Tor [5] provides anonymous communication by Onion Routing. Onion Routing ensures anonymity by using multiple proxies and encrypting communications multiple times. The network of Tor proxies does not stop even if some of the proxies go down. There are researches on anonymous information distribution using Tor [7], [8].

Some other researches also proposed anonymous communication or information distribution [9], [10], but they entirely pursue anonymity to defend privacy or rights such as the freedom of expression and the right to know. Although our proposed system also aims to protect them, it allows to trace the publisher of an article in specific cases.

Netnews [6], a traditional information distribution system, constructs a P2P network of news servers deployed by users. The proposed system also constructs a P2P network, and published articles are relayed by node to node like Netnews. However, Netnews does not consider anonymity of users. An article in Netnews contains the Path header field, which indicates the route on which the article is relayed since it had published [11]. The route is a clue to trace the publisher of the article. Our proposed system allows a user to publish an article with hiding information about the publisher of the article.

B. Public Key Infrastructure for Distributed Systems

Our proposed system requires users to have their certificates. However, the system does not use Certificate Authority (CA) to retrieve certificates since it could be an SPOF. Some researches proposed how to construct Public Key Infrastructure (PKI) as a distributed system.

Pretty Good Privacy (PGP), which is used in email encryption and digital signature adopts Web of Trust (WoT) for PKI [12]. In WoT, if you trust a user, you can trust other users who the user trusts as well. Figure 2 shows an example of a trust network of WoT. An arrow in the figure indicates that the nodes at both sides of it trust each other. For example, node $N_2$ can trust node $N_4$ through node $N_5$ and node $N_6$. A user can trust other users indirectly by WoT even if there is not direct trust relationship.

There have been some researches on distributed PKI for online publishing certificates. Chord-PKI [13] proposed a method for publishing and managing certificates using Chord [14], which is an implementation of Distributed Hash Table (DHT). In a Chord-PKI system, a system administrator deploys some trusted nodes which publish certificates into the P2P network. The trusted nodes publish certificates on demand by other untrusted nodes. Published certificates and Certificate Revocation List (CRL) are managed with Chord. There are some other researches about building distributed PKI [15]. They aim to build scalable distributed PKI in P2P networks.

Our proposed system should not be operated by a specific administrator, and hence the system adopts PGP WoT for publishing certificates.

III. THE DESIGN OF THE SYSTEM

The proposed system is an article distribution system like Netnews. It provides only the following simple two functions for publishing articles to a user:

1) Publishing articles to other users

2) Reading articles published by other users

The goal of the proposed system is to protect the privacy of users from censorship and surveillance, while the system allows a user to trace the publisher of an article as well. The requirements for the proposed system are defined as (1) availability, (2) pseudonymity, and (3) traceability. In addition, the proposed system provides a method to detect impersonation described as (4) impersonation-proofness. The rest of this chapter explains them.

A. Availability

When unfavorable information for a government is published, the government may attack to stop the system or block access to the system so as to suppress the spread of the information. The proposed system adopts P2P architecture, so as to contain any SPOF, attack to which might cause system down. In addition, if a system were operated by a specific organization, a government could conduct surveillance by intervening in it. The proposed system does not require a specific administrator.

Figure 3 shows the overview of the system. The system constructs a P2P network of users’ terminals. A user can...
deploy a node to the network and publish or retrieve articles through the node. A published article is relayed by node to node and spreads through the network. In this architecture, if each node connects to multiple nodes, a published article will spread through the network even though some nodes stop. The system therefore has higher availability than a system based on client-server architecture with respect to not containing any SPOF. Additionally, the network is composed of terminals of users, and no specific administrator is needed. This means that the system does not have the same problem as the surveillance of the NSA mentioned in I.

If some nodes in the network are attacked and are made to stop, the network may become partially disconnected. However, a node can relay articles to other connected nodes as long as they are alive even when some parts of the network are disconnected. The network of the system therefore does not always require to be a fully connected one.

A node peers with another trustworthy node. As mentioned in Section II-B, each node has their own certificates. A node can determine whether it can trust another node or not by verifying its certificate. If a node can trust the certificate of another node, it can peer with the node. The proposed system adopts PGP WoT for PKI. Trustiness of a certificate can be examined with searching the WoT network.

B. Pseudonymity

A user in the system can publish articles pseudonymously. Pseudonymity is a concept different from anonymity. General meaning of anonymity and pseudonymity in terms of information distribution is described as follows.

Anonymity
A user can publish information, hiding his identity.

Pseudonymity
A user can publish information using an alternative identifier, hiding the real identity. It can be interpreted that you can detect the user’s identity by gathering some pieces of information.

In this paper, the term “pseudonymous” means that a user can publish an article with hiding his identity but other users can detect the identity of the publisher of the article by gathering some pieces of information. This section explains how a user publishes an article with hiding his identity.

A node which has received an article relays it to its neighbor nodes. Figure 4 shows how an article is published and relayed to neighbor nodes.

In the figure, a user publishes an article through his node $N_x$. In this paper a node which receives an article from its owner is called an initial node, and a node which relays an article from another node is called a relay node. $N_x$ relays the article to node $N_y$ as if $N_x$ were a relay node, so that other nodes including $N_y$ cannot find the distribution path of the article. An article does not contain any information about who published it. This means that the user has published the article with hiding his identity. Thus the proposed system enables pseudonymous publishing by multiple relaying.

C. Traceability

The previous section has described how a user publishes articles with hiding his identity. This section discusses how a user traces the initial node of an article by gathering some pieces of information. An initial node receives an article from the publisher of it, and in many cases the node is deployed by the publisher himself. Tracing an initial node can be a clue to trace the publisher.

Users in the system are supposed to have their own certificates. As discussed in Section II-B, the proposed system adopts PGP WoT for PKI, which does not require a specific administrator. A user can find information contained in a PGP certificate by tracing WoT. A PGP certificate generally contains the name, email addresses, etc., of the owner. The authenticity of these information relies on processes of signing certificates. In the common use case of PGP, users sign each other’s certificates with face-to-face communication. Some certificates might be associated with information about a fake owner, but association can be more reliable by using some formal documents such as a driver’s license which contains the owner’s name and a face shot. The proposed system uses PGP certificates since they are already used by many users. It is important that users easily join in the system. If a fake certificate having a signature of another user is found while tracing process, the trace stops at the user who signed the certificate.

The rest of this section explains how a user traces the initial node of an article. In the system, a user records a relaying log in his local storage when he relays an article from a neighbor node.

The format of a relaying log is as follows.

```
log := (sender_id, article_id, previous_log_hash, signature)
```

- **sender_id**: The id of the node from which a node received an article.
- **article_id**: The id of an article explained in Section III-D.
- **previous_log_hash**: The hash value of the log stored in the source node of relaying which was recorded when the node received the article from one of its neighbor nodes. If
the source node of relaying does not have a relaying log for some reasons (in many cases, it is the initial node of the article), this value can be a random value. This value is passed from the source node.

signature
A signature of the relaying log. This is added by the source node of relaying.

If each node records a relaying log for each relaying, a user can trace the initial node of an article by gathering logs. Now suppose all relaying logs about an article are gathered from each node. Each relaying log contains information about from whom a node received the article (sender_id) and the hash value of the log which the node whose id is the sender_id has (previous_log_hash), so a tree like Figure 5 can be constructed by joining the information. The P2P network on which articles are relayed generally has cycles like Figure 3, but joining relaying logs generates a tree owing to previous_log_hash. The root of the tree is the initial node of an article, so a user can trace the initial node if he gathers all relaying logs about the article.

However, a user can easily tamper with logs since they are stored at local storage, or can show fake logs in order to pretend that he relayed the article. To address this problem, the source node of relay signs a relaying log which the destination node records. The key used for signing relaying logs is the secret key which corresponds to the certificate of the source node. It means that the destination node outputs a relaying log for some reasons (in many cases, it is the initial node of the article), this value can be a random value. This value is passed from the source node.

A user can examine whether a node is the initial node of an article or not with a relaying log which the node has. If the node can show a relaying log about the article, it is not the initial node but a relay node. Conversely, if it cannot show a log, it is the initial node. If it did not receive the article from the publisher and just relayed from another node, it should have a relaying log signed by the node from which it has received the article. An initial node cannot obtain such relaying logs. If a node cannot show a signed relaying log about an article, it is regarded as the initial node of the article nevertheless it just have relayed the article from another node. This means that a relay node which cannot show its relaying log may be regarded as an initial node. A node must not relay an article if the node from which it received the article does not sign its relaying log.

A user can trace the initial node of an article by gathering relaying logs. If all users along the path from the initial node show their relaying logs, the user can trace the node. Each user determines whether to show his log or not considering he wants to trace the initial node. If he wants to trace the node, he should show his relaying log. If he does not want to, he should not. If some of the users hide their logs, the user cannot trace the initial node.

D. Impersonation-Proofness

Users in the system publish articles pseudonymously, and hence a malicious user can publish articles impersonating another user. This is not a main issue this paper focuses, but such impersonation confuses other users. The system allows a user to attach a proof that the publisher of an article is the same one of another article which the user has published (but other users cannot find who is the publisher) to an article.

The format of an article is as follows.

\[
\text{article} := (\text{article_id}, \text{body}, \text{publickey}, \text{signature})
\]

article_id
The article’s id. This is defined as the hash value of body and publickey, so this value changes if even one of them is modified and the article will be regarded as different one.

body
The body of an article.

publickey
An public key generated by the publisher. This is a different key from the public key used for the certificate of the publisher. The owner’s information is not associated with the key unlike certificates, so other users cannot obtain information about the publisher from publickey.

signature
The encrypted value of article_id with the secret key corresponding to publickey.

A user generates a pair of a secret key and a public key to construct articles. It has been explained that a user has a certificate and signs relaying logs with the secret key corresponding to the certificate, but the key pair used to construct articles is different one from them. It means that a user has two key pairs for different purposes. Note that no information about the key owner is associated with the key pair used for articles.

If a user uses a single key pair for multiple articles, other users can verify the articles are published by the same user. When a user obtains two articles, the user can examine that by verifying the signature of one with the publickey of another. If the verification succeeds, the user can be convinced that they are published by the same user. Users in the system thus detect impersonation using public key cryptography.

IV. Evaluation

This chapter discusses several features of the proposed system.

A. Traceability

The proposed system introduces a method for tracing the initial node of an article by cooperation of users. Cooperation means that a user shows his relaying log if he wants to trace
the initial node. The more users show relaying logs, the easier they trace the initial node. In order to evaluate how users’ actions affect traceability, suppose that a user can take one of the following three kinds of actions about showing relaying logs.

1) Not show a relaying log: a user does not want to trace the initial node
2) Not show a relaying log until traced to him: a user does not positively want to trace the initial node but cooperates in tracing it if logs are traced to him
3) Show a relaying log: a user wants to trace the initial node

An article is relayed by node to node, and hence a tree is constructed by gathering relaying logs. Now we suppose each node relays an article to n nodes respectively. In that case, a tree of relaying logs is an n-ary tree like Figure 7.

Let \( N_d \) be a node whose depth from the root is \( d \), and \( T(N_d) \) to be the probability that relaying logs are traced to \( N_d \). Relaying logs are traced to \( N_d \) when either of the following cases occurs.

1) At least one child node of \( N_d \) takes the action of showing a relaying log
2) Relaying logs are traced to at least one child node of \( N_d \) and the child node takes the action of not showing a relaying log until traced to him

Let \( x \) to be the probability that a user does not show a relaying log and \( y \) to be the probability that a user does not show a relaying log until traced to him, where \( 0 \leq x, y \) and \( 0 \leq x + y \leq 1 \). The probability that a user shows a relaying log is \( 1 - (x + y) \). \( T(N_d) \) is described as follows.

\[
T(N_d) = 1 - \prod_{i=1}^{n} \left( x + y(1 - T(N_{d+1})) \right)
\]

Let \( h \) to be the height of a tree, and a leaf node is represented as \( N_h \). The probability that relaying logs are traced to a leaf node is defined as \( T(N_h) = 0 \). \( T(N_{h-1}), T(N_{h-2}), \ldots, T(N_{h-m}) \) are determined as follows.

\[
T(N_{h-1}) = 1 - \prod_{i=1}^{n} \left( x + y(1 - T(N_i)) \right)
\]
\[
= 1 - (x + y^n)
\]

\[
T(N_{h-2}) = 1 - \prod_{i=1}^{n} \left( x + y(1 - T(N_{h-1})) \right)
\]

An initial node is represented as \( N_0 \), and hence the probability that relaying logs are traced to an initial node, that is, \( T(N_0) \) is determined as follows.

\[
T(N_0) = 1 - (x + y(x+y)^n)^h
\]

Figure 8 shows a plot of \( T(N_0) \) for each \( y \) where \( n = 3 \) and \( h = 10 \). The vertical axis represents \( T(N_0) \) and the horizontal one represents \( x \). \( T(N_0) \) is a monotone decreasing function of \( x \) and \( y \). It indicates that the more users take actions of not showing a relaying log or not showing a relaying log until traced to him, the more difficult it is to trace the initial node of an article. Conversely, it is easier to trace the initial node if more users show relaying logs.

B. Government Intervention

The purpose of the proposed system is to provide users with opportunity of free expression. However, if unfavorable articles for a government spread over the system, the government may intervene in the system to suppress the spread of the articles or trace the publishers of them. This section discusses attack to break the availability and pseudonymity of the system by a government.

1) Attack to Availability: A government may attack the system to suppress the spread of unfavorable articles. Several methods to attack the availability of the system are conceivable.

Attack to the network

A government can make a system stop by attacking SPOF if the system contains it. The proposed system is designed as a distributed system not containing any SPOF, and hence articles continue spreading even if some nodes in the system stop. The network of the system may become partially disconnected when some nodes suddenly go down. However, the system does not required to be a fully connected
network at all times since a node can relay articles to another node as long as they are alive.

Attack to PKI

The proposed system requires users to have their own certificates. Users cannot join in the system if a government attacks the PKI used in the system and stops it. As explained in Section II-B, the proposed system adopts PGP Web of Trust for publishing certificates. PGP Web of Trust does not require a central CA, and hence it is difficult to stop it.

Deploying many nodes to impede the spread of articles

If there are nodes that do not relay an article, the spread of the article is impeded. A government can deploy many nodes that do not relay articles into the network. However, users in the system can choose nodes to peer with, and even if some nodes stop relaying, an article can spread from other nodes which have it. It is difficult to completely stop the spread of an article by this method.

2) Attack to Pseudonymity: The proposed system allows a user to publish an article pseudonymously. This section discusses attack to the pseudonymity.

Deploying many nodes to trace the publisher

In the system, a user traces the initial node of an article by tracing relaying logs. A government can trace the initial node more easily by deploying nodes near the initial node. In order to facilitate tracing arbitrary nodes, a government can deploy many nodes into the network and shorten the distances from other nodes. Users in the system have their own certificates and can choose nodes to peer with, and hence they should not peer with nodes which seem to be spies of a government.

Enforcing showing relaying logs

In order to trace the initial node of an article, a government may enforce showing relaying logs on users. In that case, most users would show their relaying logs. However, government’s authority generally affects only people within the borders of the nation and not affects foreign nations. In order to investigate a foreign user, a government needs consensus with the government of the country on investigating. Governments would cooperate in investigating users if serious articles such as a plan of terrorism had published, but investigation of criticisms to a government could not get consensus with other governments. Thus traceability of cross-border cases is affected by consensus among governments, so eliminating authoritarian censorship by a single government, which is one of the purposes of the proposed system, will be achieved. Users therefore should peer with nodes in foreign countries as much as possible, referring to their certificates.

V. Conclusion

This paper has proposed a P2P article distribution system considering both users’ privacy and criminal investigation. In the system, if many users prioritize privacy, their anonymity is protected, but if many users prioritize criminal investigation, they can trace the publisher of an article. In order to realize that, each node in the system records a relaying log in its local storage when it relays an article. In evaluation, traceability and cases of government intervention have been discussed.

There are several issues for future work. (1) We should examine the proposed system on a large scale environment. (2) The proposed system needs a function to recommend nodes to peer with to a user. (3) We should consider enhancing the reliability of a certificate. (4) The efficiency of article distribution should be improved.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number 26280030.

REFERENCES


