Introduction

of

Anonymous P2P Communications Platform

backtrack routing for dynamic robustness

matching & mining of sensitive data under anonymity

for

Secure Information Infrastructure

May 7, 2003

Tatsuya Kainuma
Senior Research Engineer
Fujitsu Prime Software Technologies Limited
Aoi 1-16-38, Higashi-ku, Nagoya, Japan
E-mail: kainuma@pst.fujitsu.com

Anonymous Peer-to-Peer Communications Platform

backtrack routing for dynamic robustness

matching & mining of sensitive data under anonymity

Data-mining of sensitivel information must come with the protection of privacy. In alliance with Prof. Hideki Imai of University of Tokyo and Mnemonic Security, Inc., we have developed the prototype of a communications platform that satisfies this objective by having the user ' s identity verified reliably while the anonymity is securely assured for communication, data storage and matching/mining.

1. Background

1-1. Minimization of Damage

We should note the critical nature of leakage of privacy information.

Restoration is just impossible.
Descendants could be influenced.
Punishment of culprits makes no solution for victims.

We should also note that there can be no such a thing as an unbreakable data center in terms of leakage of information, since it is not just the system or a building around it, but men of flesh and blood who work there that leaks or helps leak the information.

We assume that sensitive data that could leak will leak. The issue is, then, how to minimize the damage when leakage of privacy data and other sensitive information has occurred. The best that we can expect and hope to achieve is to minimize the damage to the smallest unit, that is, one individual. The solution that we propose is to depend less on centralized management of personal information and more on the distributed management built on a P2P network.

1-2. Personal Data Mining

Suppose there is a person who is occasionally a visitor to an online bookshop. His purchase record would make the target of the bookshop ' s data mining operation and he would receive various offers and suggestions, which he would probably find to be just a nuisance. He visits the online bookshop only when he did not find the books he wanted at real bookshops in his neighborhood. Data-mining of such piecemeal purchase record could be only spreading the antipathy to the vendors among their otherwise faithful customers.

Next, assume that we have received a pertinent proposition from an online shop where we have made few purchases. Would you be pleased or would you feel frightened? We might well mutter " I have not exposed all myself to them. How have they learned so much about me? How much of my private information have they gathered? Can I be certain that they will not leak it, purposely or accidentally? If they do leak my personal information, whether purposely or accidentally, should we not punish them for letting it happen? "

For online shops or data centers, it is a nightmare of leaking privacy. For consumers, it is a nightmare of having privacy leaked.

1-3. Striking a balance

We need to strike a balance. The system that we develop contains the following principles.

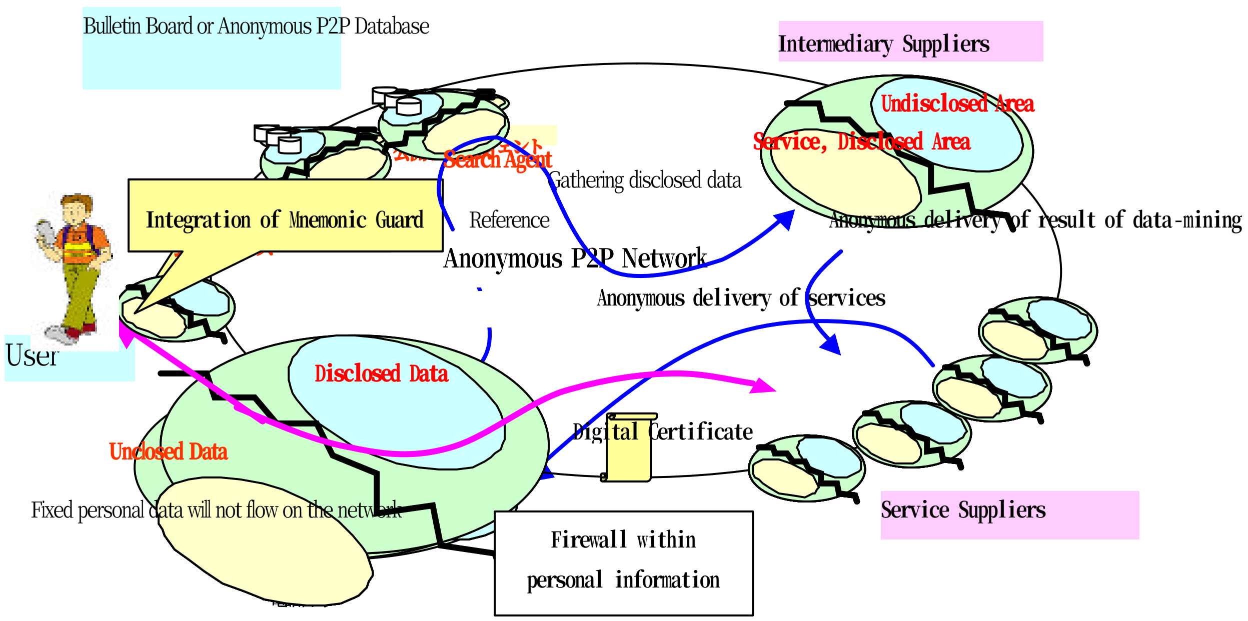
Self-responsibility: Personal identification data should be managed by individuals, only by the individuals.

Maximizing security: When privacy is to be handled, we should secure the anonymity of the individuals on the network, and distributed database would be preferable to centralized ones for the sake of robustness.

Meaningful effects: Only under anonymity, we could assuredly provide our comprehensive personal data to the suppliers of services and products for expecting meaningful results out of data matching/mining.

Assurance & Usability: Should a third person be allowed to tamper with our health data stored on the network, we could be killed. Anonymity on the network makes sense only when it comes with the secure personal verification. And the user verification must be as stress-free as possible so that every one of us including the elderly can easily practice it.

2. Standard Model



User identification data, such as name and address, and otherhighly secret information will not go into the network, but contained in users ’ devices. Data to be matched/mined or the reference thereof will be dispatched to the bulletin board or distributed over the network through more than a few onion routers for securing the anonymity. Intermediary suppliers could collect the data and process them in such a way that the service suppliers can easily get a meaningful result out of the data matching/mining. The service suppliers send the proposals to the anonymous users, say, without knowing the real identity of the users.

Users could turn into services suppliers or intermediary suppliers, and vice versa.

2-1. Reliable Personal Verification

Secure online user verification by Mnemonic Guard helps protect the data stored at the bulletin board or anonymous P2P database from being tampered by other people.

A user should be able to opt to abandon the anonymity at any time in order to get a fully personalized service. Without such a free choice, the anonymous communication would perhaps be a castle in the air. Disclosure of the identity would, however, cause a huge problem on the side of service suppliers, who would then be held responsible for managing and protecting the personal information, which is no longer just the object data for matching/mining, but exactly the privacy. The supplier should obtain from the user a letter of intent for protection, and the user should also obtain a letter of intent from the supplier. The letters of intent should be accompanied by a digital certificate, which should be issueed as a result of volitional brain work, say, Mnemonic Guard personal verification in our model.

2-2. Dynamic Onion Routing with Backtracking Functions

Onion routing is the key technology for securing anonymity. For P2P platform, we should be able to make the routing management dynamically, since we have no ways of knowing which peers will stay in the network for long and which peers may get out without previous warning. Being able to make retrials by backtracking is indispensable. When onion routing fails, we could resort to multicasting. All of those trials should be done under anonymity in that any peer can send the message to any peer without knowing their real identity. A third person who is monitoring the traffic should have no way of knowing what message is traveling from whom to whom.

3. Outline of Development

The first prototype model that we developed was for a healthcare service. The function of this system is that blood pressure, weight and other health data are taken from every user terminal and uploaded to the health data management center anonymously, and statistical and analytical results from these collected health data are sent back to the users. Individual users are supposed to classify their attribute information into publicized information and secrete information, the latter of which are stored only in the users own information device (PC, PDA etc.). The model has the following properties:

* The network connecting individuals and service providers (medical agencies and its sub-agencies) is constructed by P2P (peer-to-peer) to avoid system weaknessof a management server unpredictably breaking down.

* Publicized information of individuals is shared in P2P network anonymously (with identifier) and made available for data mining by service providers. XSR (XML Service Registry), a proprietary object database technique owned by Fujitsu PST, is used in this information sharing mechanism.

4. Structure

Applications were developed for anonymous P2P network platform with anonymous BBS (bulletin board = public database), service uses and service providers and intermediary suppliers operated on it, and with Mnemonic Guard as the online personal verification technology.

The anonymous P2P network platform is implemented with the following properties:

(i) Nodes located in the middle of the communication route cannot know the origin or the final destination of the data.
(ii) In emergency the legitimate body of multiple entities should be able to remove the anonymity when the request is made by the legitimate authority (like a group of doctors) in accordance with the prior agreement.
(iii) Proxies (Onion Routers) are free to participate or withdraw by their own judgment, and clients that use them should be able to seek the proxies dynamically.
(iv) A reply can be sent back to the sender anonymously.

5. Software specification

(1) Functions and terminology

a) Anonymous network and its platform

The anonymous network platform is realized by onion routing of peer nodes during data transmission. Onion routing is one of the methods that routing peer nodes (and network skimmers) can only be aware of a part of the route information while the multi-encrypted route information will be retained till the destination of data packet. Anonymous P2P network platform differs from ordinary onion routing, as the name indicates, in that P2P onion routing is applied in an environment where peer nodes are supposed to change dynamically. The P2P onion routing enables any peer to seek communication routes anonymously, adapting dynamically during routing to structure change caused by unpredicted departure of peer nodes.

* P2P onion router : function unit to process P2P onion routing at peer nodes
* P2P onion route : multi- encrypted route information that can cope with structure changes
* P2P onion route seeking protocol : a protocol to find P2P onion route to the destination peer.
* P2P onion proxy : function unit at peer nodes of seeking P2P onion route protocol

b) Authentication of connection to the platform

Anonymous communication platform connection authentication is the function operated on a newcomer node when it enters this platform by connecting to the first node.

The first node is assumed to have disclosed its IP address, port number and so on in advance. The P2P onion route and other cryptographic schemes used in the anonymous communication platform makes use of RSA cryptosystem. The RSA public key, its certificate and secret key of every entity are supposed to have been registered and held by every peer in advance. This public key certificate serves as a passport to the community made of this anonymous communication platform, authenticating the node when it tries to enter the anonymous communication platform.

(i) The first connection node (s) is determined by a preference file configured in advance.
(ii) The P2P onion proxy of a newcomer node begins connecting to the adjacent node(s) by forwarding its public key and certificate.
(iii) The P2P onion proxy of the first connected node then passes its own public key and certificate back to the newcomer node.
(iv) Both nodes verify the certificates. Communication would be stopped if either one of the certificates

is invalid.

(v) Both nodes exchange data encrypted by the public key of the counterpart. If the data are returned correctly after decryption, the newcomer node is judged to be a legitimate peer on the anonymous P2P network platform.

(c) Anonymous route seeking

The anonymous route seeking is the function of anonymously collecting the P2P route information from sender peer to receiver peer. The route information is made of peer ID, public key and node information that constitutes the route to the receiver node. The route is chosen out of the multiple candidate routes from node information. The P2P onion route seeking protocol builds the P2P onion route in the following way:

(i) Every peer will broadcast the route seeking packets to find a route on the P2P network when booting afresh when it has connected to the first node or when communication has been disconnected due to errors.

(ii) A peer in the middle of the route having received the broadcast using P2P onion proxy will determine the returning route from its own node to the sender node and attach this information to the route seeking packet and forward all the packets.

(iii) Once the destination node has received the route seeking packet, it will return the route information packet to the sender node through the P2P onion route configured on the way to the destination. When multiple numbers of packets have reached the destination node, multiple numbers of route information packets will be returned to the sender node.

(iv) The P2P onion proxy at a peer in the middle will attach its own node ID, public key, and node information to the packets and forward these packets towards the sender node. The information added afresh will be encrypted with the temporary public key that had been set in the route seeking packet by the sender node.

(v) At the end of the sender node, the P2P onion proxy will pick up the optimal route according to the node information obtained from various route information packets and construct the P2P onion route.

(d) Anonymous forwarding / returning function

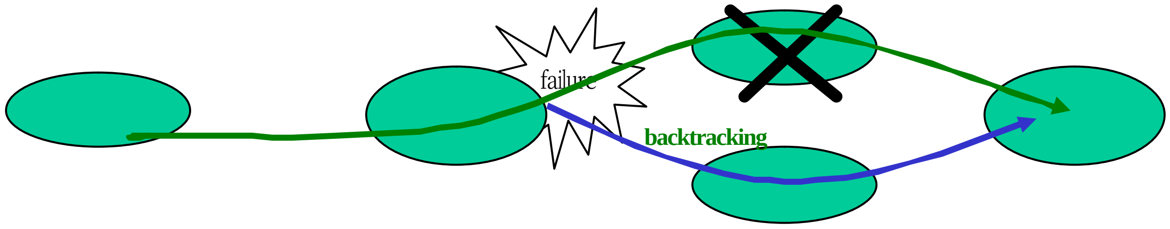
Anonymous forwarding / returning is the function that a P2P onion router transmits the data anonymously from/to the sender peer to/from the receiver peer referring to the P2P onion route contained in the data packet. The P2P onion router works in the following manner:

(i) The P2P onion route will be decrypted with the secret key of the node in work. The data will then be transmitted to the next P2P onion router via the route determined as the optimum of all the routes that had been sought. The node will also re-encrypt the group of other unselected candidate routes (Backtrack onion) as well as the return route with its public key, re-set the P2P onion route and pass the packet to the next relay node.

(ii) The receiver peer, while retaining its anonymity, can return the received message to the anonymous sender peer using the return route information stored in the onion route which had been constructed during the P2P onion routing. The P2P onion routing in returning has only one P2P onion route. Except that, it is the same as the one for forwarding.

(iv) Backtracking: The data packet will reach the destination by repeating the onion routing, but the repetition could be disturbed on a P2P network due to the structural change and other reasons. When the data fails to reach the next node during transmission, the best of the next candidate nodes will be selected from backtrack onion for the retrial, which could be repeated backwards.

(v) P2P Multicasting: Should the last trial of the backtracking fail after trying all the possible routes, multicasting will be used as the last resort. Even in this P2P multicasting, the returning route will be attached to the P2P onion route in the message transmission to the next node, just like the onion routing explained above. When a message transmission failure occurs in the returning P2P onion routing, P2P multicast will also be used to ensure the delivery of the returning messages.



(e) Message unlinkability function:

One-time anonymous IDs will be used for peer nodes for the enhanced unlinkability of the messages. The anonymous ID will function to identify the returning messages in the P2P multicast, too.

(i) The anonymous ID will be generated from a Hash value and a random number. Only the very node that owns the secret key can verify whether the anonymous ID belongs to itself. The hash value will be generated from a PriKeyY + a Random Number.

Anonymous ID	
Hash value	random number

The PriKeyY is the secret key, one of the unique pair keys that are obtained from a certificate authority that issues member passports. When a message is returned by multicast to the node of its own anonymous node ID, the receiving node can determine whether this message belongs to it or not in the following way:

* An anonymous ID is given to the message submitted to the BBS.

* When receiving a multicast message, the node will generate a Hash value from the PriKeyY and the random number contained in its own anonymous ID. If this value is the same as the one contained in the anonymous ID, the message will be verified to be the one addressed to this node.

(ii) If this anonymous communications platform is used for a community wherethe anonymity has to be removed in emergency, there could be such an arrangement that the anonymous ID (including a random number) will be sent to the Certificate Authority, where all the secret keys will be tried to generate all the possible anonymous IDs, and find one which is identical to the one in question.

(2) Anonymous BBS (public database)

a) BBS

Anonymous BBS (public database) has the following functions:

(i) Submission of a message
(ii) Search of the submitted data.
(iii) Response to the submitted data.

The features are as follows:

*It is not possible to tell who has posted or is posting a message.
* It is not possible to tell who has searched or is searching the submitted data.
* Submitted messages will be automatically deleted after a certain period of time.
* When a searcher finds no target data, it is possible to make a search reservation so that the searcher will be notified when the target data have been posted.
* The searching reservation will be automatically cancelled after a certain period of time.

b) Tamper-resistance and confidentiality

It is impossible for anyone else to tell the author of a message from the ID because the one-time ID is used for anonymity. A message holds the public key of the sender, which enables the replier to encrypt the reply message. For preventing the message receiver and network skimmers from find the identity of the sender of a message, public keys of peer nodes are managed in the following way:

The sender node generates a pair of temporary public and secret keys. The temporary secret key will be encrypted by the public key of its member passport, which will then be attached to the message together with the temporary public key. The receiver will encrypt the reply message by the sender ’ s temporary public key and send it back to the original sender. The reply message can be decrypted only by the sender node that has the secret key of the member passport

1) Generation of the message to be submitted

Key pairs needed for the messages for submission to BBS should be generated in fixed data format.

Key type	Public key	Secret Key
Fixed key pair for a user	PubKeyY	PriKeyY
Fixed key pair for BBS	PubKeyQ	PriKeyQ
Temporary key pair	PubKeyX	PriKeyX

Submission data

Anonymous ID		Message posted	Public key for reply	Secret key to decrypt the reply
Hash value	Random number	PubKeyQ (Messaga to be posted)	PubKeyX	PubKeyY(PriKeyX)

* Anonymous ID

To be made of a pair of Hash value and random number, with the hash value to be generated from PriKeyY +Random number)

* Message posted
To be encrypted by PubKeyQ.
* Reply message
To be encrypted by PubKeyX.
* Secret key PriKeyX to decrypt the reply message
To be encrypted by PubKeyY

2) Decryption of the posted message at BBS
PriKeyQ will be used for decryption of the message received by the BBS.

3) Generation and transmission of reply message
Below are the keys used for generating the reply message.

Anonymous ID		Reply message	Secret key to decrypt the reply message
Hash value	Random number	PubKeyX(Reply message)	PubKeyY(PriKeyX)

* Anonymous ID
To remain the same as that of message posted.
* Reply message
To be encrypted by PubKeyX.
* Secret key to decrypt the reply message
To be the secret key for the reply message, PubKeyY(PriKeyX).

4) Decryption of the reply message
The secret key PriKeyX should be used to decrypt the encrypted reply. However, since this secret key is encrypted by PubKeyY, the PubKeyX should first be decrypted by PriKeyY. Should the decryption by PriKeyY be successful, the PriKeyX will come out. The reply message can be decrypted by the PriKeyX.

6. Applications and Markets

The platform can incorporate cellular phones besides PDAs and PCs as the terminal devices for theanonymous communication, so it will eventually have the markets for such extensive applications as follows:

Medical/Health
Emergency response
Public services
Financial services
Recruiting services
Exhibitions, expositions and theme parks
One-to-one electronics commerce
and
Wherever anonymous communication is needed for matching/mining of sensitive data.

Furthermore, this anonymous P2P communications platform may well be part of the Internet of the next generation, on which the maximum degree of information assurance must be offered to all the participants.

May 7, 2003