A Protocol of Unlinkable Transaction for Preserving Customer Privacy*

Seok-kyu Kang † Tomoyuki Asano ‡ Kwangjo Kim †

Abstract—In this paper, we consider the protocol which prevents the service provider from finding out which customer have bought what kind of contents by the unlinkability between the payment and user-profile information. Besides, we do not employ any kind of anonymous payment system causing more computation complexities and overheads to the network, and our approach can be easily applied into the current implemented payment mechanisms. While customer’s privacy is being protected, customers are required to reveal their identities to pay for desired contents and the service provider is able to get the necessary information for its marketing activities. To achieve described above, we design the RSA blind signature-based system architecture that protects the customer privacy for the digital content transaction.

Keywords: privacy, e-transaction, blind digital signature, anonymous communication

1 Introduction

Over the past year, as the number of Internet user has been tremendously getting increase, various business and technologies based on the Internet were developed. Above all, the e-commerce is the biggest market in these days but the concern of privacy protection is getting influential. The service provider automatically collects customers’ information to analyze and learn their purchasing patterns and inclination for the personalized advertisement and maintaining the customer relationship. However, the user who accesses sensitive web sites or wants to remain hidden on the network is dissatisfied with these kinds of personalized services. Moreover, it is possible that if companies deal illegally their customer information with other companies without any permissions or the customer information leak out of companies accidentally occurred, the privacy would be infringed and broken. We believe that the privacy protection features provide business advantages to the service provider. If two service providers sell the same digital contents with the same price while one of them provides privacy protection and the other does not, the former is definitely more attractive to customers. To provide useful privacy protection, many proposed protocols and applications have been introduced so far, but they are only useful for web surfing in which users have no desire or not required to be identified. However, when customers wish to make online purchases using their credit card numbers or banking accounts, they need to provide some identifying or authenticating information. In such situations the issue of privacy is not user anonymity communication problem, but how to hide customers’ shopping/surfing patterns as much as possible. In general, the explicit demands from customers and businesses regarding commerce based on the consumers’s personal information are as follows. Customers want to be able to control their privacy perfectly while shopping over the network. The company needs reliable and various customer information and an access channel to analyze customers’ purchasing pattern or dynamic market trends. This problem is essentially conflicting to the anonymity communications problem. The former is concerned with hiding user’s surfing activities from the server but the user is required to reveal identification information to the server while latter is concerned with hiding user’s identity but all the user’s surfing activities are under the prey eyes of the server.

The rest of the paper is organized as follows: In Section 2, we review anonymity communication and previous works in the literature [1][2][3][4][5][6][7]. In Section 3, we describe some requirements that the system should be provided and the architecture of our proposed system. In Section 4, we evaluate the security and performance of our scheme. Finally, we compare our scheme with other works in Section 5 and conclude in Section 6.

2 Background

In this section, we introduce basic concepts which would be used in our proposed scheme and previous works for preserving privacy over the network.

2.1 Anonymity Communication Channels

During the past years, several kinds of anonymity-preserving network systems have been proposed and some kinds of systems are classified as multi-proxy based systems. These systems, such as Crowds, Onion-routing, etc, employ a number of network nodes between user and web server. In these systems, a series of proxies communicating over encrypted channels cooperate to forward data to a destination. The user connection anonymity is protected by the cooperation of each node on the network. We will briefly explain functions and
structures of these systems in this section. Crowds[1][2] is a network infrastructure with multiple nodes and based on the idea that people can be anonymous when they blend into a crowd. This is operated by grouping users into a geographically diverse group. If any individual node receives the request, it has to randomly decide whether to submit the request or to forward it to another node again. All communication between nodes is encrypted by a shared key, thus the Crowds protects against the eavesdroppers and message attack. Onion Routing[5][6] provides anonymous connections using different layers of encryption. It operates dynamically the connection building within a network of real time Mix-network. It is hard to track packets because they can be drop out and initiated at any node. When using the Onion routing, senders choose a sequence of routing nodes and open connections by sending layered encrypted data called “onion” to the first of them. Each onion router uses its public key to decrypt one layer of encryption, pads the embedded onion to maintain a fixed data size, and forwards the encrypted remainder of onion to the next node, but the data looks different to each router because of the layered public-key cryptography.

2.2 Previous works

Recently, there have been several proposal on the privacy-preserving systems for transactions on the Internet. In this section, we discuss some previous works related with our system and suggest their weakness.

In [4], Bao and Deng concerned about the anonymous transaction and commensurate with the general problem of the Private Information Retrieval (PIR). They introduced the system that allows a customer to disclose identity information to the web site in exchange for a digital item, but prevents the web site from learning which specific item a customer intends to obtain. The potential customer pay for the desired content on the Internet but also his purchasing information is hidden from the web site. Therefore, it is difficult to get the necessary sales information for service provider’s business activities. To do this, the merchant (the web site is equivalent to the merchant) generates the secure package including item information, encrypted item and encrypted encryption key. The customer downloads this secure package with free of charge through the Internet. So to get an encryption key for item decryption, the user must obtain the key from the transaction server which is independent entity in the system architecture. Of course, the static number of downloaded item can be gathered at the service provider but such numbers cannot precisely reflect the number of sold copies of each digital item. In the real world, the sales information is very important to run business and the royalty payment.

Gritzalis, Moulos and Kostis[3] introduced the system based on the informediaries. The informediary (I/M) is a business entity supporting the development of anonymous business models and its basic role is to accumulate user information, and deals with products and services on behalf of them. In other words, the informediaries provide the user information that is not enough to identify each user to suppliers and maximize the value of customers profile while they prevent suppliers or commercial web servers from collecting user profile. Therefore, the use of I/M enables customers to increase their bargain capability without revealing personal data and, at the same time, enables vendors to promote products and service without violating customers’ privacy.

In the work[7] by Enzman, Kunz and Schneider, the proposed system prevents the vendor (or supplier) from linking the user information which is gathered while searching with identifying information. In order to do this, the system requires asymmetric algorithm for data encryption by using public key of the vendor. If the user wants to buy some products or services, the user generates agents which contain the desired product information, and sends it to the base station which is in the middle of communication between a user and vendors. That is, all agents from the user are sent to the destined vendor via the base station. The base station dispatches these agents and plays a role of proxy. Thus, the vendor cannot gather users’ IP addresses and cookies for linking the received order.

In [3][7], these approaches generate the pseudonyms for customers and employ the TTP between the customer and the service provider. The service provider is able to get the necessary information related with its customers, but the customer must trust the TTP.

3 Our Proposed Scheme

We consider new privacy preserving system for digital content transactions. In our proposed scheme, we use the blind signature and multi-proxy based anonymity channels for providing unlinkable transactions.

We categorized the customer’s information into the payment information and user-profile information. The payment information that includes the credit card number or banking account which is usually used to authorize the customer for the payment when he/she purchases any desired digital content, and the payment information is only revealed to a payment entity. (In this paper, the payment information of customer is equivalent to identity.) The user-profile information contains the user-untraceable data, such as customer’s age, gender or habit, and is used in the service provider to obtain necessary information for its business activities.

3.1 Overall Architecture

Our proposed system consists of three components: Customer, Service Provider (SP) and Payment Server (PS). The customer purchases digital contents from SP on the Internet, and has to pay for desired contents validly by revealing their information required in the payment process. SP provides digital contents to the customer and this participant should have user-profile information for its marketing activities. However, SP is not permitted to get customers’ identities, thus, it has to be difficult for SPs to trace a customer who bought a certain digital item. PS is an entity performing the payment process for the customer. In the real world, PS could be a credit-card company, a bank or payment gateway. In our approach, PS cannot know about the customer’s purchasing information, such as what the customer bought or which content he/she expects to buy. Simply, PS only deals with the payment information received from a customer and there is no way for PS to know which customer intends to purchase what kinds of contents. PS does not reject SP’s request to transfer money after finishing the transaction of the customer.
In addition, our approach uses the multi-proxy based anonymity network system, as we mentioned above, between the customer and SP. We assume that it is vulnerable to eavesdropping but robust against the traffic analysis attack.

3.2 System Requirements

We consider that the protocol is operated under the Public Key Infrastructure and use RSA-based blind signature in our proposed system. Since RSA algorithm is already, generally implemented and used in Internet browsers such as MS Internet Explorer and Netscape, our proposed system can be easily adapted in current existing network systems. Moreover, other various websites are also applying RSA algorithm for security thus these technological trends are one of the reason why we adopted RSA-based blind signature. Every participated entities have their own public key from the Certificate Authority (CA). As described in the previous section, the customer uses the anonymity network to be hide his information, such as IP addresses, cookies, in the browsing step, because if the anonymity network does not used in the browsing step, SP is able to know who have accessed and analyze the accessing patterns from staying time in its site.

I. Decentralization of customer information
The user identities and his profile information should be managed in PS and SP repeatedly. If the single system component manages and stores all information about customers, the information exposure could be more fatal than it could be at the system that disperses its customer information over participated components. Besides, customers are required to trust a single system component for being anonymous. Thus, our proposed system should scatter and decentralize customer information.

II. Providing customer identities for the payment
For applying the current payment technology and designing the system more practically, the protocol should not use the anonymous payment system. Consequently, to pay for desired digital contents, the customer needs to open his payment information but not any information related with desired contents.

III. Controlling the profile information exposure
The customer should provide his profile information selectively. That is, the customer needs to determine his preferences to be revealed to control his privacy from SP, and our proposed system needs to ensure that the providing user preferences must depend on the customer’s willingness.

IV. Unlinkability between customer identities and their profiles
Any entities except the customer itself cannot link the customer identifying information to his profile information. Even if SP and PS collude and share their information mutually, it must be difficult to find any relation between identities and profiles.

3.3 Protocol

In this section we describe our proposed system protocol, and how the unlinkability of customers' identities and their transactions are provided. Overall protocol is consist of three phases: Setup, Purchasing and Delivery. We will use the following notations to describe the protocol.

<table>
<thead>
<tr>
<th>CI</th>
<th>Content’s sample Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAYINFO</td>
<td>user Payment Information containing identification, Credit card number or account information etc.</td>
</tr>
<tr>
<td>P</td>
<td>Content’s Price information</td>
</tr>
<tr>
<td>CID</td>
<td>Individual CI’s Identification</td>
</tr>
<tr>
<td>SID</td>
<td>SP’s Identification</td>
</tr>
<tr>
<td>REFinQ</td>
<td>user REFerence inQuery</td>
</tr>
<tr>
<td>REFanS</td>
<td>user REFerence anSwer</td>
</tr>
<tr>
<td>E(), D()</td>
<td>Symmetric encryption/decryption algorithms</td>
</tr>
<tr>
<td>BlindX(K,M,r)</td>
<td>Blinding function by entity X. It accepts a public key K, a message M and secret random number r, and generates a blinded output.</td>
</tr>
<tr>
<td>UnblindX(M,r)</td>
<td>Unblinding function by entity X. It accepts a secret random number r and a message M, and generates a unblinded output.</td>
</tr>
<tr>
<td>N_x</td>
<td>Nonce of entity X</td>
</tr>
<tr>
<td>M</td>
<td>full digital contents</td>
</tr>
<tr>
<td>K,K^{-1}</td>
<td>public/secret key</td>
</tr>
</tbody>
</table>

3.4 Setup phase

Setting up Parameters  SP and PS individually set up the system wide RSA parameters as follows:

1. Pick a 1024-bit RSA module \(n = pq\) with primes \(p = 2p' + 1\) and \(q = 2q' + 1\) where \(p'\) and \(q'\) are also primes.
2. Choose a random 120-bit number \(d_p\) and let \(d_q = d_q + 2\).
3. Compute the RSA secret exponent \(d\) by the Chinese Remainder Theorem (CRT) such that \(d = d_p \mod 2p'\) and \(d = d_q \mod 2q'\).
4. Compute the RSA public exponent \(e\) such that \(ed = 1 \mod 2p'q'\).

The public key \((e,n)\) is made public and the private key \(d\) is distributed securely. We define the public/secret key for SP as \(K_{sp}\) and \(K_{sp}^{-1}\) respectively. Similarly, the public/secert key for PS are defined as \(K_{ps}\) and \(K_{ps}^{-1}\).

CI Creation  Before a transaction with customers, SP creates the content’s sample information (CI). That can be a movie trailer, a part of music file or any kind of attractable information. Note that SP creates many CIs that introduce the same digital content, thus, each CI has its own identifier (CID) to be used when SP sends a full content to the customer at the delivery phase. After SP creates CI with CID, SP stores CID into the database.

REFinQ Creation  SP also creates the REFinQ which is a questionnaire of kind and required for SP to understand customers’ preferences. To do this, SP makes questions asking the user-untraceable information, for example, it may ask about age, gender, favorites, motivation of buying and so on. It means that customers...
can specify what information should be disclosed to whom, when it should be disclosed, and for what purpose, and that they are guaranteed the information will be treated so. Furthermore, SP can analyze unspecified individual customer’s preference without knowing his/her identifier for business.

**Step.1** SP carries out the following: generate the signature for its own ID (\(SID\)) and price of i-th digital content’s \(P_i\) with a private key. After that, generate the bundle \([CI_i,\{SID,P_i\}_{KSP},\text{REFinQ}]\).

Any latent customers can download this bundle without charge via an anonymous communication network. This bundle is only one-time downloaded in order to avoid the content re-delivery.

**Step.2** After downloading the bundle, SP stores \(CID_i\) into the database because even though the digital content is not purchased yet, those information about stored CID_s could be a good statistical data for analyzing and auditing digital contents.

### 3.5 Purchasing Phase

**Step.3** The potential customer who downloaded a bundle and decided to purchase makes out the \(REFinQ\) and we call it as \(REFanS\) after answering. To answer \(REFinQ\) is not a mandatory in our system and this step totally depends on the customer’s intention. In other words, the customer does not need to make a \(REFanS\) or he can select any questions he just wants to answer.

![Figure 1: Setup and Purchasing Phases](image)

**Step.4** The customer generates his payment information(PAYINFO) to pay for desired content. The PAYINFO, in this protocol, is a credit card number, banking account, or any other information to be used for the payment process. We consider that the PAYINFO can be used as user identification and is only opened to the PS. To get a PS’s blind signature, the customer creates a random secret integer \(r\), computes \(U=\text{Blind}_{CUS}(K_{PS},CI_i,r)\) with the public key of PS. The value \(U\) is the blinded \(CI\) and the customer sends it to PS for obtaining its signature.

**Step.5** The customer encrypts a set of \(\{SID,P_i\}_{KSP}\), PAYINFO and \(U\) with the public key of PS, and sends it to PS.

**Step.6** The PS verifies \(\{SID,P_i\}_{KSP}\) whether it is generated in SP. If valid, the PS starts to process the customer’s payment request with PAYINFO and \(P_i\).

**Step.7** If the payment process is successfully processed, PS computes where \(SI=U \cdot P_i\) and compares \(P_i\) with \(P_i'=CI_i^{-1}(CI_i \cdot P_i)\) to validate whether the customer paid accurate price for his desired content, and check the state of \(CID_i\) from the database to avoid the double delivery of digital content. Since the state of \(CID_i\) is automatically changed when the matching digital content is purchased or downloaded, we can easily obstruct the double use of \(CI_i\).

**Step.7-1** PS sends \(S\) to the customer.

**Step.8** At this step, the customer receives the value \(S\). So, PS’s signature, \(S\), justifies whether the customer paid properly without revealing what the customer intends to buy. To unblind the value \(S\), the customer computes \(\{CI_i \cdot P_i\}_{KSP} = \text{Unblind}_{CUS}(S,r)\)

### 3.6 Delivery Phase

When SP delivers the digital contents to the customer, an anonymity network is employed between them. An anonymity network to be employed in our system is used to provide the connection anonymity for the customer. Thus, the customer would be anonymous from SP while he purchases the digital contents. As described before, we use the concept of multi-proxy based system to avoid that the single proxy determines the customer’s identity.

![Figure 2: Delivery Phase](image)

**Step.9** In this step, the customer demands SP to send the full digital content (\(M\)) by sending an encrypted data \(\{CI_i \cdot P_i\}_{KSP}, CI_i, \text{REFanS},N_{CUS}\) where \(N_{CUS}\) is a nonce generated by the customer.

**Step.10** SP verifies \(\{CI_i \cdot P_i\}_{KSP}\) and compares \(P_i\) with \(P_i'=CI_i^{-1}(CI_i \cdot P_i)\) to validate whether the customer paid accurate price for his desired content, and check the state of \(CID_i\) from the database to avoid the double delivery of digital content. Since the state of \(CID_i\) is automatically changed when the matching digital content is purchased or downloaded, we can easily obstruct the double use of \(\{CI_i \cdot P_i\}_{KSP}\).

**Step.11** After doing all confirmation, SP encrypts the full digital content \(M\) with \(N_{CUS}\) which is from the customer. That is, \(Z_s = E(M_i,N_{CUS})\)

**Step.11-1** SP sends encrypted full digital content \(Z_s\) through the anonymous communication channel. When finishing the content delivery, the service provider changes the state of \(CID_i\) to purchased condition in the database.
If the customer receives the encrypted full digital content, $Z_i$, he computes $M = D(Z_i, N_{CUS})$ to get his purchased digital content.

4 Analysis of the Protocol

In our protocol, since the customer blinds his desired $CI$, and $PS$ merely generates a signature for blinded $CI$, $PS$ cannot learn which content the customer intends to purchase. Even though $PS$ has customers’ identity information from $PAYINFO$, it does not provide enough information to infer or track the customers’ buying pattern because the $CI$ is blinded by the customer’s secret random number $r$ using RSA blind signature scheme. $SP$ has the customers’ profile information from the REFanS received from customers at the purchasing phase. Each REFanS contains the individual customer’s preference such as his age, gender, date of purchasing, his favorites and so on, but these preference information do not say about any customer identifying information.

In addition, all communication of $SP$ and the customer is achieved on the anonymity network so that $SP$ cannot trace the specific customer’s identity. $SP$ just delivers contents to the customer through the anonymity network after authorizing whether the customer has paid or not.

4.1 Performance Analysis

The most heavy computational burden to $PS$ is the verifying operation $\{CI_i, P_i\}_{K_{PS}}$ in step 10. The operation $\{SID, P_i\}_{K_{PS}}$ is also expensive, but it is conducted only once for each digital content $M$, while verifying operation is performed per transaction. Hence we want to reduce the cost of $\{CI_i, P_i\}_{K_{PS}}$ as much as possible. This is the reason why we choose the secret key $K_{PS}$ through applying the CRT. Since $d_p$ and $d_q$ for generating secret keys are small 120-bit numbers, the computation is much cheaper than an direct 1024-bit RSA algorithm. Since the secret key is chosen in a special way, the $e$ has negligible probability to be small.

The most expensive computation for the customer is $Blind_{CUS}(K_{PS}, CI, r)$. But this step can be done in advance as a pre-computation, i.e., the selection of $r$ and the computation of blinding function can be carried out as soon as the customer’s machine is power on or during the machine idle time. The task for the customer’s machine to do after unblinding of $S$ is getting random nonce $N$ and decrypting $M = D(Z_i, N_{CUS})$ which are cheap operations.

4.2 Security Analysis

The problem of speeding up RSA encryption algorithm has been studied in cryptography for many years. It has been noticed that choosing small secret exponent $d$ could be dangerous[8][10]. So far the best way is to choose small $d_p$ and $d_q$. The meet-in-the-middle attack with Fast Fourier Transform technique provides an algorithm of complexity $O((\sqrt{d_p} + \sqrt{d_q})^2)$ to factorize $n$[9]. Therefore, a 120-bit $d_p$ can provide a security level higher than 272, which is much lower than the cost of the best factorization of 1024-bit $n$.

One possible attack is that the collusion of $SP$ and $PS$ by sharing their information: customer identities and profiles. In our proposed protocol, if two participants share their information mutually, they cannot link customers’ identities to their profile information together since the $PS$ does not know $CI$ that the customer paid for unless knowing the customer’s secret number $r$. Also, $SP$ does not know who have bought its digital content, and there is no way to find any relevance between them.

Our approach provides protection against unfair activities by either $SP$ or the customer. Possible disputations and cheatings are addressed. All cases require the authority as a mediator. In addition, if required, the customer can reveal his identity. At the end of the transaction, even if the customer paid for his desired content to $PS$, it is possible that $SP$ refuses to deliver a full content. In this case, the customer shows his $S$ to the mediator to prove that he has already paid $P_i$ for $CI_i$ to $PS$. If valid, the customer prevails.

Another possible fraud is that the customer who downloaded a number of bundles and accumulated multiple $\{(SID, P_i)^{K_{PS}}\}$ can cheat $SP$ by replacing $\{(SID, P_i)^{K_{PS}}\}$ associated with a high cost item with another $\{(SID, P_i)^{K_{PS}}\}$ associated with a low cost item. Because $PS$ receives the $\{(SID, P_i)^{K_{PS}}\}$ and blinded $CI$ of $P_i-1$, the malicious customer can pay lower cost and request $CI$ of $P_i-1$ to $SP$ by showing the signature $\{CI_{i-1}, P_i^j\}_{K_{PS}}$ of $PS$. However, $SP$ computes $P_i' = CI_{i-1}(CI_{i-1}, P_i^j)$ and compares with $P_i-1$ corresponding with $CI_{i-1}$ at the delivery phase, and if $P_i'$ and $P_i-1$ are not the same, $SP$ easily becomes aware of the customer’s cheating.

5 Comparison

In the next comparison table, the customer privacy issues are itemized into 4 factors: Privacy control, Technical trace, Sales audit and Unlinkability of transactions.

Privacy control is that how much customers can manage or control identity information themselves. The system [4] requires customers to disclose their identities for the payment operation to the transaction server. That is, the customer provide his/her identities even though their surfing activities are hidden. In [3], the disclosure of customer profile information depends on the system proxy which has a capability to build customers’ identity information. The system [7] using agents has also high privacy control. The customer himself generates the agent and sets the attributes that will be disclosed to web servers. In our proposed system, the customer is free to answer questions in RefinQ. In other word, the customer control the degree of the profile disclosure by himself. Thus, the customer of our system has high privacy control.

Technical Traceability is the possibility of tracking customers through the technical factors, such as IP addresses and cookies. Gritzalis et al. [3] have low possibility because they play as a role of filter in between the user and the web server, so these traceable factors could be removed. In our system, each transaction between the SP and customer is occurred through the anonymity communication channel. Therefore, all technical traceability are filtered out during the transmitting on the channel.

Sales Audit is that whether service provider are able to know the sales figure, (e.g., a number of sold items, what kind of group person have bought and so on) or not. The system [4] does not provide sales figure information to the service provider. Merely, the static number of downloaded items are gathered and these
numbers are not useful in real world and unpractical. In other systems [3] their proxies provide sales information to the service provider instead of the customer. In our system, the REFans generated by customers are provided to the SP through the anonymity channel and the number of sold item can be easily recognized at the delivery phase, hence SP can get enough sales information without interfering customer identities.

Unlinkability between transactions is one of important properties in our proposed system. Bao and Deng’s scheme [4] doesn’t provide unlinkable transaction. The systems preserving customer privacy control all information on their customer as a TTP. Therefore, it is possible to expose customer identities if they are conspired together. However, although our proposed system request customers to reveal their identities like these systems, the unlinkability between user identities and purchasing history are provided by applying the RSA blind signature scheme. Especially, we modified this scheme by signing blinded content’s CI with its price information at the PS. Even if the customer information is not revealed to SP and PS but also they are colluded together, due to the unlinkable transaction property, finding the relevance between occurred transactions and user profiles would be difficult.

6 Conclusion

Throughout this paper, we have studied on privacy preserving unlinkable protocol for digital content transactions. For the concrete design, we reviewed previous related works and pointed out their problems. And then we have suggested the improved protocol with RSA blind signature scheme.

Many kinds of customer privacy-preserving systems use pseudonyms or TTP that hides all customer information from service providers, but our proposed protocol, in this thesis, uses a current implemented payment architecture instead of an anonymous payment system. In addition, customer privacy is to protected without TTP. The customer only sends PS his/her payment information as an identity in order to pay for desired content and SP performs the verification based on RSA blind signature scheme whether the customer paid validly. Since the communication between SP and the customer is achieved on anonymity network, SP cannot learn and track a content that the customer intended to purchase. Moreover, even if SP and PS may collude together and share their customer information mutually, two entities cannot find out customer’s purchasing record due to the difficulty for linking customer identity to profile information. It is commonly recognized that one of the most important issues for e-commerce of digital contents is content protection and management. This is on-going effort in a number of industrial initiatives. However, additional efforts are required to study detailed integration issues with specific content protection and management systems. Therefore, it is necessary to study how to seamlessly integrate our system with a digital content protection system.

References


