2011

k Anonymous Private Query Based on Blind Signature and Oblivious Transfer

Russell Paulet
Victoria University

Golam Kaosar
Victoria University

Xun Yi
Victoria University

Originally published in the Proceedings of the 2nd International Cyber Resilience Conference, Edith Cowan University, Perth Western Australia, 1st - 2nd August 2011

K-ANONYMOUS PRIVATE QUERY BASED ON BLIND SIGNATURE AND OBLIVIOUS TRANSFER

Russell Paulet, Md. Golam Kaosar and Xun Yi
School of Engineering and Science
Victoria University, Melbourne, Australia

russell.paulet@live.vu.edu.au

Abstract

In this paper, we consider a scenario where there are a group of clients and a database server, and a client wishes to query the database, but does not want to reveal her or his query to the server. Current solutions for this problem are based on oblivious transfer, which usually requires high communication overhead. To reduce the communication overhead, we propose three k-anonymous private query protocols. Our first protocol is based on blind signature, where the server cannot determine the identity of the querying client from the group. Our second protocol is based on k-anonymous oblivious transfer, where the server cannot tell which record the querying client wants from k records. Our third protocol is a combination of the first and second protocols. Our protocols can achieve k-anonymity and are practical in many real-life applications.

Keywords

private query, k-anonymity, oblivious transfer, blind signature, database security.

INTRODUCTION

Querying databases for information is commonplace in modern society. The information varies from online shopping queries to email queries. It is so common that most people do not even realize they are performing a query when they perform their daily routines. However, executing these queries exposes privacy concerns, as the data owner can learn what content that the client finds interesting.

One such example was when AOL released query data about its users for research purposes (McCullagh, 2006). With this data, it was possible for patterns to be extracted about the users. It was also possible for third parties to deanonymize the users, by linking them with other data sources, and learn who exactly made the queries. Hence, it has become increasingly important to develop methods to perform efficient private queries that minimise the overhead required.

A private query is where a client can request some information from a server, without the server learning what was requested. This was made possible due to the introduction of the oblivious transfer, which was conceived by Rabin (Rabin, 1981). The scheme was between two parties, Alice and Bob. Bob has a message m and wants to send it to Alice with the knowledge that she will learn m with probability of $\frac{1}{2}$. The transfer is based on the ability to compute square roots modulo a composite number.

This was extended by Even et al. (Even, Goldreich, & Lempel, 1985) to a 1-out-of-2 oblivious transfer, where the receiver (Alice), could make a choice from the sender's (Bob) two available messages. This was followed by many 1-out-of-n and k-out-of-n oblivious transfer protocols (Aiello, Ishai, & Reingold, 2001; Naor & Pinkas, 2001; Tzeng, 2006; Naor & Pinkas, 2000; Naor & Pinkas, 1999).

Oblivious transfer was followed by k out of n adaptive oblivious transfer (Naor & Pinkas, 1999; Camenisch, Neven, & shelat, 2007). The protocol allows a receiver (Alice) to get k messages one at a time, instead of all at once as with the classical oblivious transfer. This has numerous applications, which include searching for content in a database.

The oblivious transfer is considered as a stronger version of private information retrieval, which was introduced by Chor et al. (Chor, Goldreich, Kushilevitz, & Sudan, 1995). This was in the information theoretic setting. This meant that it was impossible to determine the query, even with unbounded computational power. This was achieved by distributing the data to many (non-colluding) parties. This was followed by a scheme presented in the computational setting (Kushilevitz & Ostrovsky, 1997), where replication was not necessary. Most authors distinguish between regular PIR and Symmetric PIR (Aiello, Ishai, & Reingold, 2001; Gertner, Ishai, Kushilevitz, & Malkin, 1998). Symmetric PIR also protects the database from malicious users by ensuring that the users cannot obtain more than they can legitimately access.

Providing anonymity as a method to provide privacy has also been researched extensively. In the simplest case, anonymity can be achieved using blind signatures (Chaum, 1982). However, much more complicated solutions have been introduced to reflect the complex nature of anonymous access of content. Schechter et al. introduced

an anonymous system based on verifiably common secret encodings to construct a group of users to authenticate anonymously (Schechter, Parnell, & Hartemink, 1999).

Supplementing this anonymous authentication protocol, a system was introduced to allow a user to anonymously authenticate k times, which enabled a client to use a service k times before they needed to refresh (Teranishi, Furukawa, & Sako, 2004). Further protocols were introduced to enable dynamic membership and time based membership (Nguyen & Safavi-Naini, 2005; Tzeng, 2006).

The blind signature approach to provide anonymity does not provide perfect privacy, as the owner of the data can see transactions made by the client. Perfect privacy can be achieved with the oblivious transfer, but this requires the whole database to be encrypted and downloaded. This overhead can make the oblivious transfer impractical in many applications.

The concept of k-anonymity was first introduced to allow the disclosure of sensitive data for research purposes (Sweeney, 2002). The goal of k-anonymity is to prevent the identification of records in a database, such that the data records could not be distinguished from at least $k - 1$ records. This was achieved by either generalizing or suppressing data records, or a combination of both.

In this paper, we present protocols that allow a group of clients to anonymously obtain digital content, such as magazines or newspapers, from a server which owns a database of such content. We use the k-anonymity concept to reduce the communication overhead of the oblivious transfer. Protecting both the client and server is important for cyber resilience. The features for the presented protocols are as follows.

- The server is unable to associate a client to a query.
- The server is unable to correctly identify the database item that the client is interested in.
- The client is unable to obtain digital content for which they are not allowed.

BACKGROUND

In this section we discuss the basic building blocks of our protocol. These include Blind Signatures and oblivious transfer. Using these two primitives we will construct the protocols that are presented in this paper.

Blind signatures

Blind signatures can be achieved by the client adding some randomness to the message before the message is signed. Due to this randomness, the signer does not know what the message they are signing. Therefore, it is possible for a client to use this at a later time without the concern of the server determining the user of the signature.

We will describe an example of a blind signature scheme in terms of the RSA cryptosystem (Rivest, Shamir, & Adleman, 1978). The client chooses a message m for the server to sign. Using the public key e of the server, the client calculates the following according to Equation (1), where $\mathrm{r} \in \mathbb{Z}$.

$$m' = mr^e (mod\ N) \tag{1}$$

The client sends m' to the server to be signed. The server responds by calculating s' according to Equation (2). The value s' is sent to the client.

$$\mathrm{s}' = (\mathrm{m}')^{\mathrm{d}} \ (\mathrm{mod\ N}) \tag{2}$$

When the client obtains s', the signature $s$can be calculated by using Equation (3). Where r^{-1} is the modular inverse.

$$\mathrm{s} = \mathrm{s}' \cdot \mathrm{r}^{-1} \ (\mathrm{mod\ N}) \tag{3}$$

This indeed produces the signature s for message m, as shown in (4).

$$\begin{aligned} s &= s' \cdot r^{-1} \\ &= (m')^d \cdot r^{-1} \\ &= (mr^e)^d \cdot r^{-1} \\ &= m^d \cdot r^{ed} \cdot r^{-1} \\ &= m^d \ (mod\ N) \end{aligned} \tag{4}$$

Oblivious transfer

The oblivious transfer is very similar to the blind signature in the way a public key encryption scheme is used. Its main difference is that it allows for multiple messages, compared to just one. The 1-out-of-2 oblivious transfer, as introduced by Even (Even, Goldreich, & Lempel, 1985), is described as follows (using the RSA cryptosystem).

1. Alice generates an instance of the RSA cryptosystem and sends the public key e, N to Bob.
2. Alice chooses two random messages $x_0, x_1 \in \mathbb{Z}$ and sends to Bob.
3. Bob selects one of the two messages and computes $v = (x_b + k^e)\ (mod\ N)$, where $b \in \{0,1\}$, e is the public key of Alice, and $k \in \mathbb{Z}$ is chosen randomly. Bob sends v to Alice.
4. Alice computes two possible k values as $k_0 = (v - x_0)^d\ (mod\ N)$ and $k_1 = (v - x_1)^d\ (mod\ N)$, where d is the private key of Alice. Alice sends $m_0' = m_0 + k_0$ and $m_1' = m_1 + k_1$ to Bob.
5. Bob either computes $m_0 = m_0' - k$ or $m_1 = m_1' - k$ based on his choice of x_b.

Observe that we can easily transform this protocol into a 1-out-of-n oblivious transfer by changing the number of randomly chosen messages and corresponding actual messages. This primitive will be key in constructing our protocol.

OUR PRIVATE QUERY PROTOCOLS

Before we describe the whole protocol, we will describe two basic protocols. When these two basic protocols are combined, they will construct the whole protocol or hybrid protocol. The two protocols that we will describe are a k-anonymous protocol based on blind signatures and a k-anonymous protocol based on oblivious transfer.

Protocol Model

Our protocols will reside in a two-party model consisting of a client and a server. The server will own a database of digital content (newspapers/magazines) or simply known as records. Whereas the client will have a query or index that specifically refers to a single record in the database. In the general scenario, there will be many clients interacting with the server, independent of each other. We assume that when a client wants to access a service from the server, the client can establish a secure channel and no eavesdropping is possible. We also assume that the client knows what record/file they wish to download, before they execute any protocols presented in this paper.

We consider two possible attacks on our protocol, one for each party in our model. The first attack is when the server is malicious and wants to determine the client's query. The second attack is when the client is malicious and tries to access more records than what they have paid for. After we describe the protocols, we will show that the main protocol that we present will defend against both attacks.

Private Query Based on Blind Signature

The k-anonymous protocol based on blind signatures has two basic steps. First, the client obtains a blind signature, after payment, from the server. Second, the user selects the desired record from the server's database which composes the client's query. The user supplies this information, both the signature and the query, to the server and the server returns the record to the client. This is illustrated in Figure 1.

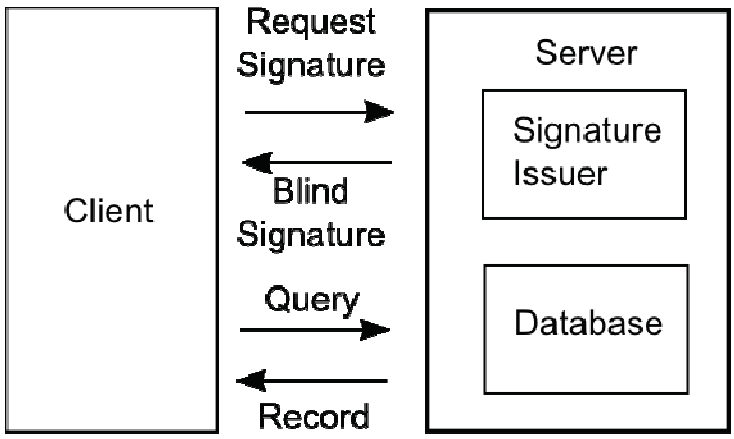

Figure 1: Overview of the blind signature based protocol.

Once the payment for the server's use is confirmed, the client proceeds with steps to obtain a signature are as follows:

1. The client chooses a random number r and random message m (both integers) and calculates m' according to Equation (1).
2. The client sends m' to the server which is signed using Equation (2), to produce the blinded signature s'.
3. The server sends s' to the client.
4. The client un-blinds by using Equation (3) to remove the blinding factor r to produce s.
5. The client can verify that the signature is correct by checking whether $s^e = m\ (mod\ N)$ or not.

Once the signature has been blindly acquired, the client proceeds with the next step, which uses the signature to download digital content from the server. We restrict that one signature downloads one file, meaning that each record is of equal price. Providing a protocol that allows for many items of different price is out of the scope of this work.

Assuming that a signature was successfully acquired from the database, the client can proceed to utilize the server's database and download a file. The steps are as follows:

1. The client initiates the protocol by supplying the signature s to the server.
2. The server verifies the signature according to $s^e = m$. If the signature is valid the protocol continues, halts otherwise.
3. The client compiles their query by choosing a record of interest.
4. The client submits this query to the server and the server transmits the record to the client.

Since the server has no knowledge of the random numbers chosen by the client, it has no way of linking a single client to a query. This is known as the unlinkability property, and is very powerful in providing anonymity. However, the server still knows what records are being accessed by the clients of the system, as they are unencrypted. This problem is solved with the introduction of the next basic protocol.

Private Query Based on k-Anonymous OT

Separate from the blind signature protocol, we describe a protocol utilizing the oblivious transfer to provide privacy. This protocol is simpler than the previous protocol as it requires only one round of communication. The steps required are as follows:

1. The server generates an instance of the RSA cryptosystem and sends the public key e, N to the client.
2. The client constructs their query by choosing a record of interest from the server's database along with other random records, totaling a subset k of the whole database, represented by the indices of the records $I_1, I_2 \ldots I_k$. The client sends this query to the server.
3. The server receives this query and generates k random messages $x_0, x_1, x_2 \ldots x_k$, and sends this to the client.
4. The client chooses one of these messages x_t, where $0 \leq t \leq k$, and computes $v = (x_t + K^e)$. Where K is chosen randomly. v is then sent to the server.

5. The server computes all possible values for K: $K_t = (v - x_t)^d \ (mod\ N)$, for $0 \leq t \leq k$, and d is the private key of the server. The values $m_0^{'} = K_0 + m_0, m_1^{'} = K_1 + m_1, m_2^{'} = K_2 + m_2 \ldots m_k^{'} = K_k + m_k$ are sent to the client.
6. Using the knowledge of the x_t chosen, uses v to calculate $m_t = m_t^{'} - K$.

At the conclusion of the interaction, the client learns his desired record and nothing additional. The server only knows that the client is interested in one of the k records, but has no way of knowing which one, due to the nature of oblivious transfer. This is illustrated in Figure 2.

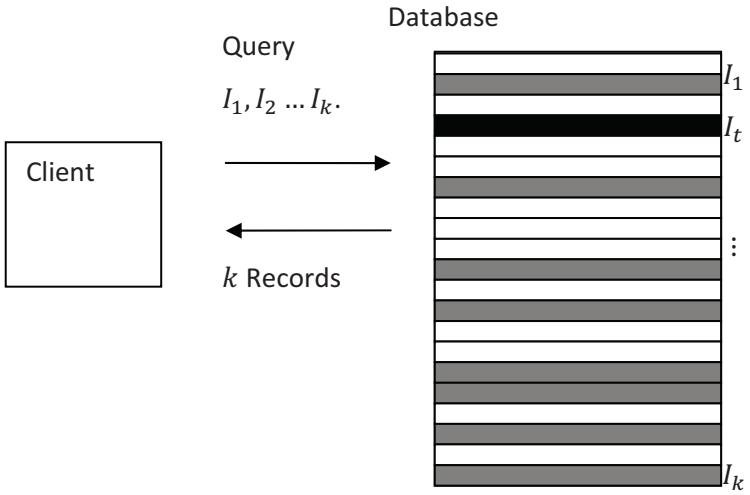

Figure 2: Displays the k-anonymity based oblivious transfer.

Remark: We suggest that the records in the database should be encrypted with a symmetric key by the server, and the client runs oblivious transfer to obtain only one key. This is because the length of the digital content we are considering will be larger than the key size.

Hybrid protocol

The hybrid protocol uses key elements from both previously described protocols, to produce a scheme that allows for a balance between privacy and efficiency. Essentially, the oblivious transfer protocol used within the blind signature protocol. The oblivious transfer is used at the point where the client is transferred the records. The whole protocol is described as follows.

1. After payment, the client requests a blind signature from the server according to our blind signature query protocol.
2. Once the client has verified the validity of the signature, the client forms a query consisting of their desired record and other randomly chosen records $I_1, I_2, I_{t,} \ldots I_k$ and sends the query to the server according to our oblivious transfer protocol.
3. The client and server engage in an oblivious transfer with the k records (subset of all records), to receive the desired record m_k.

This combines benefits of the previously discussed protocols. It both anonymises the client to prevent identification and protects the server's database against a potentially malicious client who attempts to obtain more than they are allowed. The security and performance of our hybrid protocol is analyzed next.

SECURITY AND PERFORMANCE ANALYSIS

Analyzing the security and performance is essential for any privacy preserving protocol. We will analyze these factors with respect to the hybrid protocol that we have presented.

Security Analysis

We now analyze the security of the protocol. That is, how much privacy or protection does the protocol provide the parties, the client and the server. We will analyze our hybrid protocol with respect to both the client and the server.

Privacy protection for Client

The hybrid protocol that we have presented provides two levels of privacy for the client. The blind signature component of the protocol provides anonymity for the client. Since the client is using a one-time blind signature, the server is unable to determine if two different queries belong to the same client or two different clients. This unlinkability ensures that the server cannot trace a single client.

The oblivious transfer stage of the protocol conceals the query made by the client up to a subset k, based on the concept of k-anonymity. The server knows the client is interested in one of the records from the subset, but cannot determine which record was obtained. Depending on the size of the subset, the privacy of the client's query is maintained.

The oblivious transfer stage also has an advantage over a simple anonymous based solution. If we consider this simple anonymous solution, then some information will leak through the distribution of items accessed by all of the clients. From this information, the server will be able to estimate what a single client has accessed through this distribution and potentially identify them. Figure 3 illustrates the impact of increasing k has on the distribution.

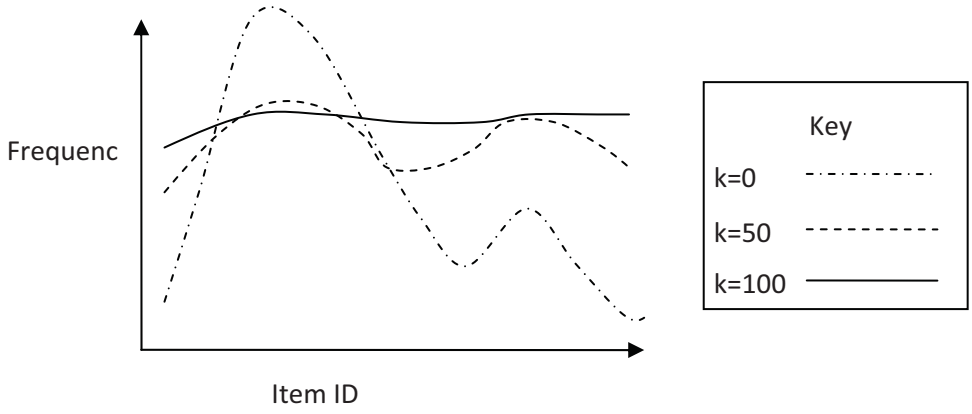

Figure 3: Hypothetical distribution of items displaying interest at two points.

As shown in the above figure, as the value of k is increased, the distribution will become more uniform. This is regardless of the original distribution. This makes it difficult for the server to make inferences about a client from this distribution.

Access Control for Server

Since the server is providing a subscription service, it must protect the records from unauthorized downloads. The server cannot simply transfer all the records to the client upon request. This will result in the loss of business. The oblivious transfer component of our protocol ensures that the client is unable to learn any more than one record, from the k records of the subset. Additionally, due to the randomness of $x_0, x_1, x_2 \ldots x_k$, the client is unable to determine other records, by trying to link multiple executions of the protocol. For instance, if a client executes two queries, then they will only be able to obtain two records. They will be unable to extract any more information from correlating the two executions.

Performance Analysis

The main performance considerations are communication and computation. We require that the protocol executes as fast as possible, without hindering the privacy benefits of the protocol. Since we are allowing the client to download a subset of the database, then we are saving resources for other purposes. We will now describe how this pertains to communication and computation.

Communication

When comparing our hybrid protocol to the classical oblivious transfer, there are dramatic savings in communication overhead. This is fundamentally because we are allowing the client to download a subset k of the whole database, while the classical oblivious transfer requires the whole encrypted database N to be downloaded. More formally, the communication complexity of records to be downloaded in our protocol is $O(k)$, whereas the communication complexity of records to be downloaded with oblivious transfer is $O(N)$.

Computation

The most expensive operation in our hybrid protocol is the exponentiation operation. It is required two times for the blind signature component, and it is required two times for each record to be downloaded. Therefore, the computation required by our protocol is $O(2k)$. The conventional oblivious transfer has no need for blind signatures, and therefore the computation required is $O(2N)$. Since k is a subset of N, k is much smaller than N and the constant in our protocol is justified. Table 1 summarizes the performance analysis of our protocol.

Table 1: SUMMARY OF ANALYSIS

	Our protocol	Oblivious transfer
Communication	$O(k)$	$O(N)$
Computation	$O(2k)$	$O(2N)$

CONCLUSION

We have presented a k-anonymity private query scheme based on oblivious transfer and blind signatures that is efficient in terms of communication and computation. This scheme achieves a balance between privacy and efficiency, by allowing the client to execute oblivious transfer on a subset of the total records, reducing the communication and computation overhead.

The main contribution of this paper is the introduction of the concept of k-Anonymity to oblivious transfer schemes. This allows the client to download a subset of records from the database, while still achieving some level of privacy. The random selection of records induces confusion into the distribution of all clients using the server's database. Hence, it is difficult to determine what a client has downloaded based on the distribution of all users. At the same time the presented protocols provide protection for the database, which is important for cyber resilience.

Future research directions include providing a framework, for which parameters for the protocol can be selected, based on the privacy and performance requirements of the application. Additionally, providing a means for the server to offer records at different prices is an interesting problem and should be investigated.

REFERENCES

Aiello, B., Ishai, Y., & Reingold, O. (2001). Priced Oblivious Transfer: How to Sell Digital Goods. *Advances in Cryptology, EUROCRYPT 2001* , 119-135.

Camenisch, J., Neven, G., & shelat, a. (2007). Simulatable Adaptive Oblivious Transfer. In M. Naor (Ed.), *Advances in Cryptology - EUROCRYPT 2007* (Vol. 4515, pp. 573-590). Springer Berlin / Heidelberg.

Chaum, D. (1982). Blind Signatures for Untraceable Payments., (pp. 199-203).

Chor, B., Goldreich, O., Kushilevitz, E., & Sudan, M. (1995). Private information retrieval., (pp. 41-50).

Even, S., Goldreich, O., & Lempel, A. (1985). A randomized protocol for signing contracts. *Commun. ACM , 28* (6), 637-647.

Gertner, Y., Ishai, Y., Kushilevitz, E., & Malkin, T. (1998). Protecting data privacy in private information retrieval schemes. (pp. 151-160). ACM.

Kushilevitz, E., & Ostrovsky, R. (1997). Replication is not needed: single database, computationally-private information retrieval., (pp. 364-373).

McCullagh, D. (2006, August 7). *AOL's disturbing glimpse into users' lives*. Retrieved 03 25, 2011, from CNET News: http://news.cnet.com/2100-1030_3-6103098.html

Naor, M., & Pinkas, B. (2000). Distributed Oblivious Transfer. In T. Okamoto (Ed.), *Advances in Cryptology, ASIACRYPT 2000* (Vol. 1976, pp. 205-219). Springer Berlin / Heidelberg.

Naor, M., & Pinkas, B. (2001). Efficient oblivious transfer protocols. (pp. 448-457). Society for Industrial and Applied Mathematics.

Naor, M., & Pinkas, B. (1999). Oblivious transfer and polynomial evaluation. (pp. 245-254). ACM.

Naor, M., & Pinkas, B. (1999). Oblivious Transfer with Adaptive Queries. In M. Wiener (Ed.), *Advances in Cryptology CRYPTO 99* (Vol. 1666, pp. 791-791). Springer Berlin / Heidelberg.

Nguyen, L., & Safavi-Naini, R. (2005). Dynamic k-Times Anonymous Authentication. In J. Ioannidis, A. Keromytis, & M. Yung (Eds.), *Applied Cryptography and Network Security* (Vol. 3531, pp. 318-333). Springer Berlin / Heidelberg.

Rabin, M. O. (1981). How To Exchange Secrets with Oblivious Transfer. *How To Exchange Secrets with Oblivious Transfer* .

Rivest, R. L., Shamir, A., & Adleman, L. (1978). A method for obtaining digital signatures and public-key cryptosystems. *Commun. ACM , 21* (2), 120-126.

Schechter, S., Parnell, T., & Hartemink, A. (1999). Anonymous Authentication of Membership in Dynamic Groups. In M. Franklin (Ed.), *Financial Cryptography* (Vol. 1648, pp. 184-195). Springer Berlin / Heidelberg.

Sweeney, L. (2002). k-anonymity: a model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst. , 10*, 557-570.

Teranishi, I., Furukawa, J., & Sako, K. (2004). k-Times Anonymous Authentication (Extended Abstract). In P. J. Lee (Ed.), *Advances in Cryptology - ASIACRYPT 2004* (Vol. 3329, pp. 81-95). Springer Berlin / Heidelberg.

Tzeng, W.-G. (2006). A secure system for data access based on anonymous authentication and time-dependent hierarchical keys. (pp. 223-230). ACM