Privacy in Location Based Services (LBS) via Composite Functions: The L4NE Protocol

Levent Ertaul
Computer Science Department, California State University, East
Hayward, CA 94542, USA

Abstract
Location Based Services (LBS) are increasingly accessed through mobile devices. This trend forced companies such as Google, Facebook, Apple, and Foursquare to provide services which incorporate location information of users. The information of an individual’s location has great significance. Today, almost all devices such as mobile wireless phones and tablets have GPS to gather the location information of their users. Despite concerns of the users about the privacy, security and third-party use of their private location information, these LBS services mushroomed everywhere. Still, even today, LBS service providers, in spite of user’s legitimate privacy concerns, are unwilling to build private LBS systems in which they don’t have access to users’ location information. The major issue in sharing location information is the level of privacy. The level of privacy in LBS is provided by using private equality testing and/or private proximity testing protocols. In these protocols, when Bob is nearer to Alice at some location within a defined proximity then the exact location of Bob is revealed. Otherwise both parties learn nothing about each other’s location. And a third party or a service provider gains no information about the users’ locations. So, to solve the privacy problem in LBS, in this paper, we present a novel security protocol, called the L4NE, based on composite functions. This protocol achieves privacy without revealing the exact location of Bob/Alice to each other unless they are in the same exact location. By doing so, they are able to achieve the privacy requirements in LBS. The proposed L4NE protocol also achieves high security with high performance. And it does not require any secret sharing, any secure key distribution protocol or any trusted/untrusted servers.

Key words:
Location Based Services (LBS), Privacy, Private Equality Testing, Private Proximity Testing.

1. Introduction

Location Based Services (LBS) are ubiquitous in today’s applications. They are an indispensable component of our communication model as LBS have proven to be crucial not only for companies but also for consumers. LBS are primarily based on user’s location information to provide other value added services by means of a wireless mobile device functioning through common cellular network or radio stations. Until now they have left a wide-ranging impact on society and businesses as a whole. Their ability to track and monitor individuals including children and its use in the law enforcement such as locating thieves, sexual predators etc. had good implications on society. A vehicle tracking device or asset tracking component for businesses, LBS technology acts as a catalyst in the growth of industries especially for telecommunication and transportation. However, as the system deals with confidential, private personal information such as location, personal mobile number and home addresses, it becomes vital for the operator to offer adequate security to maintain user’s privacy [1, 2, 3].

LBS, besides providing numerous services to consumers worldwide, are also notorious in collecting user activities. This helps them target specific products to individuals which is a market proven strategy for increased growth. Vendors of many of the mobile applications often exploit the data that is collected by the use of their services. LBS advertising -- which ties in consumer locations with restaurants, retail shops and other locations through mobile devices -- are expected to grow to over one-third of all mobile advertising in four years [2, 4, 5].

According to a study by Pyramid Research [6], LBS revenue in the US is expected to climb from $2.8 billion in 2010 to $10.3 billion in 2015. In 2015, LBS advertising revenue was around $6.2 billion. LBS advertising will generate 60% of all LBS revenue in coming years. Pyramid believes that all forms of mobile advertising will grow. “However, local search will be the most important driver of LBS advertising revenues.” Not only the developers of navigation applications will be changing their business model to fit into the local-search branch, but many different companies from different industries can also profit from the growth of the local-search market – from start-ups like Poynt and Yelp, to the local business advertising from specialized portals like the Yellow Pages, to even the search engines that are specialized for a particular topic, like Toptable or HotelBooker. The survey conducted by Pyramid shows the amount of revenue generated by the use of LBS. (see Figure 1.)
When a company monitors someone by using LBS should the concerned person’s consent be necessary? What about individual rights and personal privacy? Who is using users’ private information and how? Who is liable when personal privacy is breached in LBS? Is there any law or regulation that protects users’ privacy in LBS? These are some of the questions that need to be answered before using LBS, as monitoring, tracking, gathering, and sharing users’ private information with others may have adverse effects on some users’ lives. In the case of monitoring suspects by law enforcement agencies or security agencies, the question of individual freedom comes to mind.

To address all of the above privacy concerns, private proximity and equality testing protocols are proposed [9, 12, 15, 17, 19]. Proximity based services are a special class of LBS in which the service adaptation depends on the comparison between a given threshold value and the distance between a user and other (possibly moving) entities. The so-called “friend-finder” services are an example: Alice would like to know when her friend Bob is in the range of her proximity i.e. nearby, so that they could get in touch and possibly meet. A major privacy concern with the use of LBS is the release of the other person’s precise location information. This concern applies to privacy and equality services as well: Alice would like to use the proximity service without necessarily releasing her exact position to the service provider (SP). In some cases, she may even wish not to provide the exact location to her friends, although she may be willing to reveal whether she is in proximity or not. For example, she may agree to let Bob know that she is in a neighborhood near Bob’s location, but keep the specific address hidden from Bob. In practice, this may avoid the situation in which friends can directly talk to other friends, as the goal of the service is usually to enable communication that may only eventually lead to meetings in person [7, 8]. The proximity testing problem can be reduced to equality testing to get a better solution for privacy. For example, let $G$ be a grid of the plane and let $L_a$ be the location of user $A$, expressed as the center of the grid cell. $L_b$ be the location of user $B$. For any two users $A$ and $B$ the equality testing protocol must satisfy the following [9]

- if $L_a = L_b$ then $A$ learns $L_b$
- if $L_a \neq L_b$ then $A$ learns nothing about $L_b$ except that their location is different.

Developing privacy aware systems start with a clear definition of data sharing polices that describe the relationship between the user and service providers. Unfortunately, in today’s world, main LBS service providers like Facebook, Google, Foursquare, Apple, Samsung, SCVNGR, Yelp and Twitter etc. have built their systems on top of users’ private information with insufficient privacy preserving technologies. Nevertheless this trend seems to be changing with increasing privacy awareness and technological advances.

The research on privacy in LBS has been diverse, bringing different protocols to the solution. Most notable protocols are:

- **Synchronous Private Equality Testing (ElGamal based):** This protocol requires power calculations ($g^m \mod p$) with large numbers (minimum 2048 bits). Although this protocol is very secure, it may not be suitable for large group implementations because of the computation cost of power calculations. [9, 10, 36].

- **Fast Asynchronous Private Equality Test with Oblivious Server:** This protocol uses AES algorithm [35] for key generation. It requires a trusted secure server and secret key sharing between users and the server. In addition to that, it also requires a secret key sharing among users. This protocol does not use any cryptographic algorithms to provide privacy for location information of the users. It may not be very secure but it shows very good performance results [9, 11].

- **Homomorphic Encryptions (Pinkas Algorithm) using Location tags:** This protocol uses Pallier crypto [13] algorithm which requires huge power calculations like $(g^{r^a \mod n^2})$ with large numbers. It may not be suitable for wireless mobile devices [9, 12].

- **Relaxed Private Threshold Set Intersections:** This protocol does not use any crypto based security algorithm. There might be security issues in this protocol. On the other hand, it shows impressive performance results in Android environment [9, 14].
• **Location Privacy via Actively Secure Private Proximity Testing** (ElGamal & Schnorr based): It is computationally very expensive. It may not be suitable for mobile devices [15, 16, 36, 37].

• **A Fair and Efficient Solution to the Socialist Millionaires’ Problem** (Diffie-Hellman & Schnorr based): This protocol can be used for LBS to provide proximity. But it gets really slow when the number of users increases [17, 18, 32, 37].

• **Privacy-Aware Proximity Based Services: Hide & Seek Protocol (Based on SP-Filtering).** Hide and Seek Protocol using SP-Filtering Protocol is a simple and easy protocol which provides minimum location privacy. It is fast and it does not require crypto algorithm for privacy. This may cause some security problems. Performance depends on the Level of granulites and the number of users in the system [19, 20].

• **Privacy-Aware Proximity Based Services: Hide & Crypt Protocol (based on SP-Filtering & Commutative Encryption Functions).** Hide & Crypt is a very secure and reliable protocol. It uses a simple and highly effective encryption algorithm (stream cipher) for security. It does not reveal any information to a third party server, service provider or other users about the exact location of the service requesting user. However, the accuracy of this protocol depends on how granularity is defined [19, 21].

As discussed above, some protocols provide a very high security level with the cost of performance; some provide high performance with the cost of security. But most of them require a secret sharing among users or untrusted servers (LBS service providers) which are not desirable for most of the service users; these secret sharing and secret key distribution requirements also further complicate the privacy issues in LBS. There is a clear need to address privacy issues in LBS with proper security and performance along with no secret sharing or a key distribution problem among users and service providers.

The study in this paper attempts to address the privacy concerns of LBS by proposing a novel security protocol based on a composition of functions, with adequate security and performance, without any secret sharing or secret key distribution among users and service providers. This protocol does not even require a server for LBS services and is named as the **L4NE** protocol.

In the next section, introductory information about composite functions and a formal presentation of our new protocol **L4NE** is given which is followed by a security and performance analysis of the **L4NE** protocol. Finally conclusions are presented.

## 2. The **L4NE** Protocol

The **L4NE** protocol is a security protocol which addresses the privacy issues of LBS in equality testing by providing adequate security and performance and it is based on a composition of functions. A function, in mathematics, is a rule, which allows us to work out one set of numbers from another set of numbers [22, 23, 24, 25]. For example, by knowing the fixed cost of renting a telephone for a month, we can calculate the cost per minute to make calls. To do this, we can set up a function to work out the total cost based on the total length of the calls we have made. Combination of two or more such functions will give us a result, which is composite in nature. In general, for any two functions $f$ and $g$, the composite function $f o g$ is defined by

$$f o g (x) = g(f(x))$$

The domain of $f o g$ is the set of all numbers, $x$, in the domain of $g$ for which $g(x)$ is in the domain of $f$ [22, 23, 24, 25, 26].

The composition of functions is always associative. That is, if $f$, $g$, and $h$ are three functions with suitably chosen domains and codomains, then

$$f o (g o h) = (f o g) o h,$$

where the parentheses serve to indicate that the composition is to be performed first for the parenthesized functions. Since there is no distinction between the choices of placement of parentheses, they may be safely left off [22, 23, 24, 25, 26].

The functions $g$ and $f$ are said to commute with each other if $g o f = f o g$. In general, composition of functions will not be commutative. Commutativity is a special property, attained only by particular functions, and often in special circumstances. [22, 23, 24, 25, 26]. (This commutative property is used in the **L4NE** protocol as explained later).

Use of composite functions in cryptography is not a new concept. A couple of examples can be found in [27, 28, 29, 30].

The **L4NE** protocol uses the commutative property of the composite function. This property generally does not hold. In other words, composite functions are not commutative with some exceptions. For example

$$f o g(x) = g(f(x))$$

$$g o f(x) = f(g(x))$$
$g(f(x)) \neq f(g(x))$ generally true except if $f(x) = g(x)$

This type of commutativity is also the basic requirement in the Diffie-Hellman protocol with power functions. This commutative property holds in composite functions only by composing function $f$ or $g$ to itself. That is, if $f(x) = g(x)$ then the commutative property of composite functions will be true as shown below.

$$f(f(x)) = f(f(x))$$

or

$$g(g(x)) = g(g(x))$$

We use this commutative property in the L4NE protocol.

Another important function that we use in the L4NE protocol is

$$F(x) = F(F(F(\ldots \ldots \ldots (F(S_0)) \ldots )) \mod p)$$

where $S_0$ is initial value, $F_n(S_0)$ is $n$-composition of $F$ to itself with an argument $S_0$.

Let’s start describing the L4NE protocol as it would be used in a real system, namely at time $t$, and let’s assume that Alice is in location $L_A$ and Bob is in location $L_B$. At some certain times, Alice wants to know whether Bob is in her location or not without revealing her location to Bob. That is to say that she wants to check if $L_A = L_B$ in certain times without revealing no information about her own location.

![Fig. 2 The L4NE Protocol](image)

**Step 1:** Alice selects a large random number $N_{AB}$ and a large prime number $p$ and makes them public by sending them to Bob. Meanwhile, Alice selects another random number $n_A$ and keeps it secret as if this is her private key. Then she encodes her location information $L_A$ for that time into a function $F_A$ by choosing any method described in Section 3. And, at the end of this first step, she calculates

$$S^A = F_A^{n_A}(N_{AB}) \mod p$$

sends $S^A$ to Bob as shown in Figure 2. Don’t forget that here $F_A^{n_A}(N_{AB})$ calculation is

$$(F_A(F_A((\ldots \ldots \ldots (F_A(N_{AB})) \ldots )) \mod p)\quad n_A\text{-fold}$$

**Step 2:** On Receiving $N_{AB}$ and $p$ information, Bob encodes his location information $L_B$ for current time into $F_B$ function, selects a random number $n_B$, keeps it secret and calculates $S^B$ with $N_{AB}$ value, received from Alice as shown below.

$$S^B = F_B^{n_B}(N_{AB}) \mod p$$

and sends $S^B$ to Alice. (see Figure 2).

**Step 3:** When Alice receives $S^B$ information, she calculates,

$$S^{AB} = F_A^{n_A}(F_B^{n_B}(N_{AB})) \mod p$$

**Step 4:** When Bob receives $S^A$ information, he calculates

$$S^{BA} = F_B^{n_B}(F_A^{n_A}(N_{AB})) \mod p$$

and sends $S^{BA}$ information to Alice. (See Figure 2)

**Step 5:** Alice checks whether $S^{AB}$ and $S^{BA}$ equals to each other.

$$S^{AB} = S^{BA}$$

This equality will be true if and only if

$$F_A^{n_A}(F_B^{n_B}(N_{AB})) \mod p = F_B^{n_B}(F_A^{n_A}(N_{AB})) \mod p$$

which means Bob is in the same location with Alice. Otherwise it will yield to an inequality

$$F_A \neq F_B$$
which means that Bob is NOT in Alice’s location. This is true due to the commutative property of composite functions mentioned above.

In the next section, discussion about the security and performance analysis of the L4NE protocol is given.

3. The Security and Performance Analysis of the L4NE Protocol

The L4NE protocol has the following characteristics

- In Step 1, as shown in figure 2, the protocol requires the encoding of location information as a function. This can be done in various ways as explained by the following three methods:

  **First Method:** If the system is using location tags [9, 31] (generally this location tag information is a 32 bit integer number which identifies 2^32 possible locations) then a function in the form of a polynomial can be constructed in the following way [12]:

  \[ F(y) = (x_1y)(x_2y) \ldots (x_ky) = \sum_{u=0}^{kc} a^u y^u \]

  where \( x_1, x_2, \ldots, x_k \) are the roots of the polynomial.

  **Second Method:** A nonlinear polynomial function can be written directly from the location tags, representing binary Is with an x value with its bit power and ignoring binary 0s. For example, if the location tag’s most significant eight bit is 11100100 and least significant 8 bit is 00001011 then a nonlinear polynomial function can be written as

  \[ F(x) = x^{32} + x^{31} + x^{30} + x^{27} + \ldots + x^3 + x + 1 \]

  **Third Method:** Another way of creating functions is to encode location information as a set \( P \) of points \( \{(p_1, x_1), (p_2, x_2), \ldots, (p_n, x_n)\} \), and find a polynomial function \( F(x) \) of degree \( n-1 \) defined by the points \( P \) by using Lagrange interpolation [9, 12].

- Other functions like linear functions \( F(x) = ax \) or nonlinear \( F(x) = x^r \) or \( F(x,y) \rightarrow (y, xy) \) type feedback functions can also be used in this protocol [27]

- It is synchronous which means both parties need to be online

- Both parties perform two self-compositions

- There are two rounds, namely Alice sends a message without the existence of a server to Bob and Bob responds to Alice

Based on the characteristics of the L4NE protocol described above, let’s focus on the characteristics of the \( F(x) \) function. The \( F(x) \) function used in the L4NE protocol must have the following properties:

- Computation of \( S^A, S^B, S^{AB} \) and \( S^{BA} \) must be easy.

  \[ S^A = F_A^{nA} (N_{AB}) \mod p \]
  \[ S^B = F_B^{nB} (N_{AB}) \mod p \]
  \[ S^{AB} = F_A^{nA} (F_B^{nB}(N_{AB})) \mod p \]
  \[ S^{BA} = F_B^{nB} (F_A^{nA}(N_{AB})) \mod p \]

  where \( p \) is a large prime number around 2048 bits and \( n_A, n_B \) are large enough positive integer numbers.

Any function \( F(x) \) generated from the location coding techniques that is described above must satisfy this condition. Our preferred function is a nonlinear polynomial function from a performance and security point of view.

- Recovery of \( n_A \) and \( n_B \), by knowing \( S^A, S^B, S^{AB}, S^{BA} \), even though the function \( F(x) \) has been guessed or estimated, must be hard.

In the L4NE protocol calculations of \( S^A, S^B, S^{AB}, \) and \( S^{BA} \) can be generalized as

\[ S^N = F_x^{nN} (N_{AB}) \mod p \]

As seen from the above equation, finding \( n_N \)’s value is a very well-known discrete log problem. It means that knowing \( p, N_{AB} \) and \( S^N \) will not reveal any information about \( n_N \)’s value. In other words, finding \( n_N \)’s value from equation \( S^N \) is very difficult [32].

The selection of the function \( F(x) \) which is an encoded version of the location information, as seen from the above conditions, determines the security of the L4NE protocol. For a secure L4NE protocol, as proven in [27], any location encoding technique which generates the \( F(x) \) function must generate a nonlinear \( F(x) \) function; otherwise the L4NE protocol will not be secure. Further
research is required to determine which specific nonlinear functions will perform better for more security.

If all the above conditions are satisfied, in the \textit{L4NE} protocol, when Alice receives $S^B$ information from Bob, she will not be able to recover Bob’s location information without $n_B$’s value. The only way that Alice can find out the value of $n_B$ is to solve the discrete log problem which is very hard to solve. This makes Bob’s location information private and secure. The same thing is true for Bob. Having received $S^A$ value, he will not be able to learn Alice’s location information without the knowledge of $n_A$. In addition to that, $S^{AB}$ and $S^{BA}$ values don’t reveal any information about Alice’s and Bob’s location. At the end, Alice knows that Bob is in her location, otherwise learns nothing about Bob’s location and vice versa.

As far as third parties, malicious or not, who are very interested in learning Alice’s or Bob’s location information are concerned, they will not be able to determine both parties’ location information from publicly known $N_{AB}$, $p$, $S^A$, $S^B$ and $S^{BA}$ values without the knowledge of $n_A$ or $n_B$ which are secured based on discrete logs.

Since computing power $g^m \text{mod} p$ and product of powers such as $g^a g^b \text{mod} p$ are very expensive [33], it seems that computing

$$S^N = F_{x^nN} (N^{AB}) \text{ mod } p$$

$n_N$-self composition of $F_x(N_{AB})$ is much more inexpensive since this calculation does not require multiplications rather it depends on self-folding. This gives the \textit{L4NE} protocol a great advantage as far as performance is concerned. Further research is required on the implementation of self-composition of the $F(x)$ function to determine how faster this function calculation compared to power calculations. In the \textit{L4NE} protocol, Alice and Bob perform only two compositions. This further shows that, with a proper selection of the $F(x)$ function, the \textit{L4NE} protocol will be faster than all the above mentioned power calculations based protocols.

As described in figure 2, the \textit{L4NE} protocol does not require any secret key sharing scheme, any secure key distribution protocol, any parameter set up procedure for users or any existence of trusted / untrusted server/servers. This way, the protocol eliminates all the vulnerabilities, threats, and security weaknesses coming from these schemes, protocols, and procedures.

4. Conclusions

In this paper, the \textit{L4NE} protocol which is a new private equality testing protocol based on composite functions is presented. In this protocol, we used $n$-self composition of functions. We have shown that self-composition functions are as secure as the discrete logarithm problem if a nonlinear function is selected properly. This is the proof that the \textit{L4NE} protocol is as secure as most of the previously designed protocols if not more secure then some. It is also shown that the calculation of self-compositions is much faster than power functions. This is also proof that the \textit{L4NE} protocol’s performance is better than most of the previously designed protocols. In other words, in this paper, it is shown that a novel private equality testing protocol with better privacy and performance for LBS services based on composite functions exist. At the end of the protocol, the \textit{L4NE} protocol does not leak any location information of Alice and Bob to anybody. It keeps their location information private. That means that at the end, Alice only knows that Bob is in her location, otherwise learns no information about Bob’s location. And Bob also gets nothing about Alice’s location if he is not in the same location as Alice. In the \textit{L4NE} protocol, further research is required to find better functions which may improve the security and performance of the protocol. In the future, it is also important to address implementation issues.

References


[27] R. A. Ruepel, “Key Agreement Based on Function Composition” Eurocrypt’88, LNCS 330, pp. 3-10, 1988


[37] Schnorr C., "Efficient Identification and Signatures for Smart Cards", Proceedings of CRYPTO ’89, 1989

Dr. Levent Ertaul is a full time professor at the California State University, East Bay, USA. He received a Ph.D. degree from Sussex University, UK in 1994. He specializes in Network Security and Cyber Security. He has more than 75 refereed papers published in the Cyber Security, Network Security, Wireless Security and Cryptography areas. He also delivered more than 40 seminars and talks and participated in various panel discussions related to Cyber Security. In the last couple of years, Dr. Ertaul has given Privacy and Cyber Security speeches at US universities and several US organizations. He received 4 awards for his contributions to Network Security from WORLDCOMP. He also received a fellowship to work at the Lawrence Livermore National Laboratories (LLNL) in the Cyber Defenders program for last 5 years. He has more than 30 years of teaching experience in Security issues. He participated in several hacking competitions nationwide. His current research interests are Wireless Hacking Techniques, Wireless Security, and Security of IoTs.