PathForge:: Faithful Anonymization of Movement Data

Sebastian Kay Belle Marcel Waldvogel
Department of Computer and Information Science
University of Konstanz
Konstanz, Germany
<first>.<last>@uni-konstanz.de

Oliver Haase
Department of Computer Science
Konstanz University of Applied Sciences
Konstanz, Germany
haase@htwg-konstanz.de

ABSTRACT

For most mobile networks, providers need the current position of their users to provide efficient service. The resulting motion data is not only an invaluable source for analyzing traffic or flow patterns, but also for tracking an individual's whereabouts, even without their knowledge. Today, many carry at least one mobile networked device with them wherever they go, day and night. The resulting motion data can be used to reveal the most intimate details of our lives, making this information extremely privacy sensitive. In this paper, we present *PathForge*, a lightweight solution, which not only fulfills the provider's efficiency requirement, but continues to allow flow pattern analysis, yet provides full privacy for users when not actively involved in a call.

Categories and Subject Descriptors

C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design—*Wireless Communication*
; C.2.4 [**Computer-Communication Networks**]: Distributed Systems
; K.4.1 [**Computers and Society**]: Public Policy Issues—*Privacy*

General Terms

Algorithms, Design, Security

1. INTRODUCTION

Nowadays, we are used to carry our cell phone or similar mobile device with us all the time, though, most of us are not aware of the consequences this behaviour implies: We are continuously being tracked by our service provider. Carriers initially start tracking because of scalability and efficiency issues, which are the very reason our current cellular networks have been able to accomodate the increased proliferation of mobile devices by orders of magnitude. Whenever data needs to be routed towards a user's cell phone, carriers needs at least a very good guess about the location of the user to avoid flooding the network.

Today we are willingly – however, mostly unaware – utilizing new techniques embedded in modern cell-phones (e.g. GPS) to alleviate continuous tracking by the legal authorities. Krumm and Horvitz[2] have shown that destinations of people can be extrapolated with a high probability given only a part of their path, even if the person has never visited the destination before.

MobiHeld'09, August 17, 2009, Barcelona, Spain.
ACM 978-1-60558-444-7/09/08.

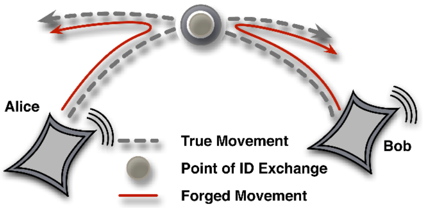


Figure 1: Path forging by switching user IDs

In this paper we propose a novel idea – we will refer to as *path forging*[1] – that addresses the following issues:: (1) Obtain accurate motion data, e.g. for traffic and flow analyses, whilst impeding malicious parties from inferring sensitive information. (2) Minimizing impact on the operator's infrastructure or efficiency, and (3) to embed the solution into a carriers' infrastructure while leaving control and implementation of the privacy-preserving system to the client systems.

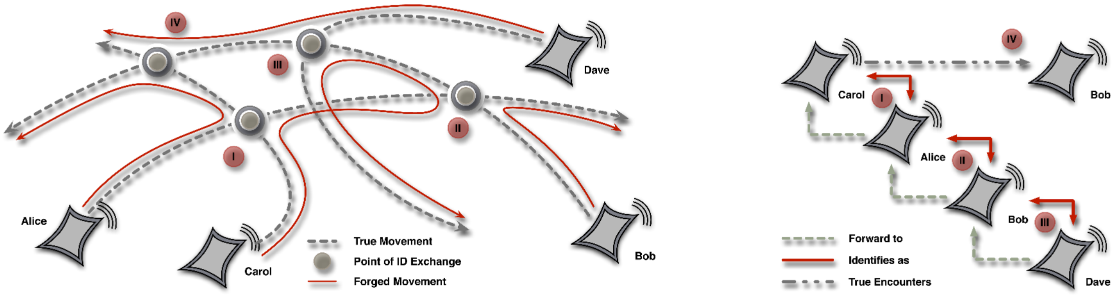
2. FORGE YOUR PATH

In *PathForge* we intend to separate user location from user identification while retaining an existing infrastructure. For a detailed description of *PathForge* and related work we refer to our technical report on *PathForge*[1].

Basic Idea:: Consider two users in the network, Alice and Bob as well as the service provider Susan. Whenever Alice and Bob meet, they swap their IDs. Each path of the phones retains full fidelity, but the coupling to their users' IDs is broken (Fig.1). Both, Alice and Bob, have a secret key K, e.g. stored on their Subscriber Identity Module (SIM), and an associated ID based on their key that is computed as e.g. $I = h(K \parallel 1)$, where $\parallel$ denotes concatenation and $h()$ a cryptographic one-way hash function. Alice and Bob use I to register with an AP and thus the network's location database. From the point of exchange Alice will identify herself as Bob, using Bob's ID (I^B) as her *Proxy ID*, and vice versa, thus, forging their motion data by utilizing the motion of each other.

Delivering incoming data to the intended person is straightforward for Susan. Whenever data needs to be transmitted to Alice, Susan will initially contact Bob, which is currently

[1] Webster's Revised Unabridged Dictionary (1913) – *Forge* (Page:585):: (2) *To form or shape out in any way; to produce; to frame; to invent*, (4) *To make falsely; to produce, as that which is untrue or not genuine; to fabricate; to counterfeit, as, a signature, or a signed document*



(a)

(b)

Figure 2: Switching IDs multiple times. The Roman numerals in (a) correspond to the numbers in (b).

posing as Alice. Thus, Bob tells Susan to forward the data to I^B. To verify that the connection is made to the authorized phone, subsequently, Susan requests Alice to authenticate using their knowledge of K. Therefore, Susan sends Alice a nonce N to compute an authentication ID as $A = h(K \parallel 2 \parallel N)$, prompting, Alice to send back A for authentication. Similarly, to initiate data transmission over the service provider's network, Alice needs to authenticate to be able to transmit the data. Note, whenever Alice or Bob initiate a connection or accept incoming data, their true identity will be revealed, and both will set back their false IDs to their initial values as computed, thus, both need to find a new partner for ID switching afterward. Due to the added overhead of transmitting calls and data messages in multiple cells, we need the users to show their real position for such high-volume transfers.

The problem we face with this initial idea is that a malicious entity (Mallory) can easily infer the true location of either Alice or Bob through backtracking.

Multiple ID Switching:: Judging from the backtracking vulnerability described above, 1×*ID switching* is not enough to prevent a malicious party to infer the location of a user. Therefore, we need to prevent the ID swapping location from being identified. This can be achieved by swapping IDs multiple times with different users along the way, shuffling identities like a deck of cards.

Consider four users in the network, Alice, Bob, Carol, and Dave. All of them can exchange their IDs multiple times with each other (Fig.2a). Therefore, Mallory will now have an upper bound of $\Omega(n^n)$ paths to infer the true location of Alice, where n is the number of encounters, though, in this case not only the direct encounters of Alice but the encounters on the path forged by I^A.

After having solved the backtracking problem, we are facing yet another puzzle:: (1) In the worst case Susan must contact every user in the look-up cascade (Fig.2b) to identify the intended receiver, (2) every user in the cascade and the network in general is a potential point of failure, and (3) every participating user in the cascade must be notified that their false IDs must be revoked after any user in the cascade revealed her true ID.

To solve this puzzle, we introduce $L = h(K \parallel 2)$ for every user. Anytime two users switch their ID, both will also exchange L along with an agreed seed s. Utilizing, s we initialize a PRNG on both client devices that computes pseudorandom time intervals r. These time intervals are subsequently used to register with the carrier as $T = h(L \parallel r)$. Anytime a user comes into a new cell in the network or T was re-generated due to a change of r, the client will register with (1) T to be reachable as a destination whose identity will only be revealed at the receipt of a call or the generation of a billable event, and (2) I to provide a forwarding point to the real identity. Thus, T cannot be used to infer any movement data and does not pose a privacy risk. Subsequently, delivering incoming data to the intended receiver will now involve T instead of I, thus, Susan ends up with only two authentication requests – the last user in the cascade and the first user in the cascade (the intended receiver). Thus, we solved two of the three previously identified problems, namely (1) contacting every user in the cascade, and (2) there is only a single point of failure left, the last user in the cascade. After Susan forwarded the data to the intended receiver, all Proxy IDs need to be revoked as they are invalid from this time on. For a detailed description of the revocation procedure we refer to our technical report [1].

3. CONCLUSION & FUTURE WORK

Our main concern in *PathForge* is to prevent malicious parties from inferring the true location of the users in the system while maintaining accuracy of motion data for traffic and flow analyses. Anonymization is solely based on switching user IDs, thus, our approach works on user level and does not need any centralized system. Additionally, even when a user reveals their true identity due to incoming or outgoing data, the revealed location information will not suffices to infer either the path the user has taken nor the destination of the user in the future as the revealed locations are only samples of the true movement pattern (c.f.[2]). A problem that still remains is the topic of anonymous data transmission and reception, removing the need for revealing a position ever.

4. REFERENCES

[1] S. K. Belle, M. Waldvogel, and O. Haase. Pathforge:: Faithfull Anonymization of Movement Data. Technical report, University of Konstanz, April 2009. http://kops.ub.uni-konstanz.de/volltexte/2009/7524/.

[2] J. Krumm and E. Horvitz. Predestination: Inferring Destinations from Partial Trajectories. In *Ubicomp*, pages 243–260, 2006.