CITY UNIVERSITY OF HONG KONG
香港城市大學

A Multi-Layered Approach to Privacy: Ensuring Privacy From the Network Level to the Application Level
由網絡層到應用層對隱私的多層保護方案

Submitted to
Department of Computer Science
電腦科學系
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
哲學博士學位

by

SCHLEGEL Roman Samuel

July 2012
二零一二年七月

Abstract

With computers becoming more powerful every year and storage space becoming cheaper, ever-larger amounts of data can be collected, analysed, correlated and stored. This has significant privacy implications for Internet users, as it becomes feasible to track, analyse and store their every move online. Users can be tracked across multiple layers using different technologies; for example, wireless access points forming a distributed WLAN can track users on the link layer, an ISP can track users on the network layer and any website can gather information about users on the application layer (e.g., using cookies, as well as using the communication content itself).

Previous work has mostly focused on individual layers and how to ensure privacy on a single layer. It turns out, however, that when using for example a system to protect the privacy of a user on the application level, a lot of information can still be leaked on other levels, for example the network layer. To better protect the privacy of users, it becomes necessary to look at inter-dependent privacy implications across several layers. In this thesis we examine the combined impact on the privacy of users by looking at two important layers, the network layer and the application layer, and propose concrete solutions on how to ensure better privacy by applying protection on different levels.

Network Layer: The threat to privacy on the network layer stems from the fact that any entity who can monitor a network link on the path between two communicating endpoints can determine that a communication is taking place, and very often also has access to the content of the communication. Depending on the nature of the communication, it can be desirable that not only the content remains private, but also the fact that a communication is taking place.

Existing solutions for anonymity on the network layer such as Tor provide good anonymity, but at the cost of usability and with significantly higher latencies. For example both Tor and JAP (a centralised anonymity system) increase the average latency by an order of magnitude from 0.4 seconds to around 4 seconds compared to direct connections. Similar concerns exist for systems such as I2P, and centralised systems (e.g., JAP) require a user to trust the operator of the system. Moreover, many of these systems only anonymise individual applications, and even then often require reconfiguration on the user side.

In this thesis we try to address these issues, and present a new approach for constructing an anonymous network on the network layer by building an overlay network on top of a conventional IP network. The overlay network decouples the actual IP addresses of nodes from virtual addresses that the nodes are using within the overlay network and the anonymity comes from the disassociation of virtual addresses from real IP addresses. By providing an anonymous network layer, the overlay also supports almost any application transparently and with minimal effort, making it more easily usable than existing systems.

In addition, an important goal is to ensure a *low latency* within the overlay compared to existing solutions, to make it more practical for daily use. To this effect, we also propose a suite of routing protocols designed to minimise latency within the overlay, preserve anonymity by not leaking topology information and also being resistant to attacks based on path cost reductions. Traditional routing protocols leak network topology information, thus compromising anonymity, while existing anonymous routing protocols do not provide authentication for routing information, allowing attackers to influence routing maliciously. We first introduce a routing protocol which does not leak topology information and we achieve this by modifying a path vector routing protocol. Specifically, we replace the path vector with a *pseudo path vector*, which detects routing loops without leaking actual route information by using encrypted random counters. Each node in the overlay can thus detect when a route loops back to itself, without having any information about the path of the route. To prevent path cost reduction attacks, we propose to modify the path cost tag in the routing protocol using cryptographic primitives such that a node can only increase the path cost. In the proposed protocol, it is easy to increase the path cost of a route, but mathematically unfeasible to decrease it, thus preventing path cost attacks by preventing a node from offering seemingly better routes and re-routing traffic to itself.

Simulation results using real-world network traffic data sets (CAIDA) show that the combination of the anonymous overlay network and authenticated anonymous routing protocol can reduce the latency by up to 50% compared to Tor and JAP.

Application Layer: Collecting information about a user on the application layer is even easier than on the network layer, as the entity collecting the data is typically explicitly involved in the communication (as compared to the network layer, where for example an ISP is usually only an observer). Considering the actual communication content to be part of the application layer, any entity involved in the communication can use the content to learn information about a user, store transmitted content and refer to previously transmitted content to build a detailed profile of a user.

On the application level we focus on privacy with respect to two specific, popular applications, namely social networks and location-sharing services.

- **Social Network Privacy:** The use of social networking sites can incur substantial privacy risks because many users disclose a significant amount of personal information. Some existing solutions to this problem such as “flyByNight” or “Facecloak” solicit an external third-party server to provide online privacy protection of content

shared by users on social networking sites. Other solutions such as "NOYB" incur a key distribution overhead among the users who are sharing content. These solutions usually also have a noticeable impact on the user experience; "flyByNight" for example integrates as a Facebook application and does not allow to protect content posted normally, making its use distinctly different from the normal usage of Facebook. Systems which use an external server shift the security trust and system reliability from the social networking provider to a third-party which still needs to be trusted and may be more susceptible to single-point-of-failure issues.

To address the shortcomings in existing solutions, we propose a new system in this thesis which can achieve the following two features through a novel application of a constant-size-ciphertext broadcast encryption scheme: (1) social networking content posted by a user can only be read by authorised users and nobody else, not even the social networking site itself; (2) no key distribution or any external server is necessary during normal operations. Apart from a key extraction server which is contacted only once by each user, our system is entirely self-contained within the web browser of each user using a plugin architecture. The system can be used directly with existing social networking sites and transparently encrypts content being sent to the a site, and decrypts it on demand when browsing encrypted content.

A thorough evaluation of a prototype implementation for the social networking site Facebook shows that the scheme is indeed feasible, scalable and practical.

- **Location Privacy:** Location-based services are privacy-sensitive because the location information necessary to provide a service allows to track a user's movements in significant detail. At the same time, such services can also be very valuable and useful. Services such as Google Latitude or Apple's "Find My Friends" allow users to locate their friends, while Foursquare and Gowalla provide "check-in" functionality to mark the places one visited. Common to all these services is that the service provider has access to all the locations submitted by users, which is a significant privacy concern.

  Existing systems designed to protect location privacy in such services either still rely on a trusted third-party with access to the location of all users, use expensive algorithms or protocols in terms of computational or communication overhead, or can only provide approximate answers to location-based queries.

  In this thesis, we improve on the privacy of these schemes and present a new system for location-based services. Specifically, we introduce two different variants, one which is applicable to point of interest (POI) searches around a user's location, and another variant which offers "friend finder" functionality, while still preserving the location privacy of individual users.

  - **POI search:** For POI applications we design a dynamic grid system that only requires a semi-trusted third party, which does not have any information about a user's location. In addition, the communication cost for the user is independent of

the user's desired privacy level, it grows linearly with the number of relevant points of interest in the vicinity of the user. The system uses two non-colluding servers (a query server and a service provider) together with a dynamic grid system and encrypted identifiers. The service provider only knows the general query area (which can be chosen arbitrarily large by the user), and the query server only sees encrypted grid identifiers which do not leak location information. Experimental results also show that our system is more efficient in both computation and communication cost than state-of-the-art privacy-preserving algorithms in many situations, with both communication cost and computation cost often lower by an order of magnitude (e.g., 0.4ms compared to 6.3ms and 0.6KB compared to 10KB for kNN queries among 10'000 POIs).

- ◇ **Friend Finder:** To enable friend finder applications with strong privacy, we design a system that allows friends to share their exact location without the need for a trusted third party or having to reveal location information to a server or users outside a group of friends. The system uses an encryption algorithm which allows to make relative distance comparisons between encrypted tuples, without revealing the actual distances involved. Through the sophisticated application of a system of encrypted location markers, friends within the vicinity of a user can be easily located. The system also achieves low communication cost by allowing users to receive the exact location of their friends without requiring any direct communication between users. It also only requires one round-trip between a user and the server to find all friends in the vicinity. Another important feature is that it provides *personalised privacy protection* within a group of friends through the use of *privacy markers*, which indicate a maximum distance at which a user is willing to be located by his/her friends. Experimental results show that the communication cost is at least 50% lower (e.g., 8.5KB compared to 17KB for groups of 1000 users) than the cost of state-of-the-art solutions depending on the parameters (number of friends, range, etc.).

We conclude that by considering privacy implication on several layers and combining schemes to protect privacy on different levels simultaneously, the total leakage of personal information can be minimised, and as a result personal privacy online can be improved.

Contents