AFRL-IF-RS-TR-2005-131
Final Technical Report
April 2005

PROTECTING THE PRIVACY OF INDIVIDUALS IN TERRORIST TRACKING APPLICATIONS

Palo Alto Research Center

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE
ROME RESEARCH SITE
ROME, NEW YORK

STINFO FINAL REPORT

This report has been reviewed by the Air Force Research Laboratory, Information Directorate, Public Affairs Office (IFOIPA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

AFRL-IF-RS-TR-2005-131 has been reviewed and is approved for publication

APPROVED: /s/

PATRICK K. MCCABE
Project Engineer

FOR THE DIRECTOR: /s/

JOSEPH CAMERA, Chief
Information & Intelligence Exploitation Division
Information Directorate

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE
APRIL 2005

3. REPORT TYPE AND DATES COVERED
Final Apr 03 – Sep 03

4. TITLE AND SUBTITLE
PROTECTING THE PRIVACY OF INDIVIDUALS IN TERRORIST TRACKING APPLICATIONS

5. FUNDING NUMBERS
C - F30602-03-C-0037
PE - 62301E
PR - GENI
TA - SY
WU - S2

6. AUTHOR(S)
Teresa Lunt, Paul Aoki, Dirk Balfanz,
Glenn Durfee, Philippe Golle, Diana Smetters,
Jessica Staddon, Jim Thornton and Tomas Uribe

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto California 94304

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory/IFEA
525 Brooks Road
Rome New York 13441-4505

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
AFRL-IF-RS-TR-2005-131

11. SUPPLEMENTARY NOTES

AFRL Project Engineer: Patrick K. McCabe/IFEA/(315) 330-3197/ Patrick.McCabe@rl.af.mil

12a. DISTRIBUTION / AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 Words)*

Countering terrorism involves gathering information from a wide diversity of sources to discover key facts and relationships, develop models of hypotheses, and support human reasoning on likely futures and outcomes. Many of these data sources contain sensitive personal information, such as data on telephone calls, email, credit card usage, bank accounts, car rentals, housing, educational data, health-related data, drivers' licenses, airline and hotel reservations, visas, border crossing, attendance at events, and application for government programs. In tracking potential terrorists and attempting to discover their relationships and organization, is it necessary to focus on data about individuals. Yet it is this identification of data with individuals that makes the information sensitive. The goal of this project was to allow authorized analysts to search these data for terrorist-related activity while providing a realistic degree of privacy protection for ordinary citizens who may be also represented in those databases. The proposed solution has the following elements: Inference control to prevent unauthorized individuals from completing queries that would allow identification of ordinary citizens, access control to return sensitive identifying data to appropriately authorized users, Immutable audit logs that ensure all data accesses are recorded immediately and permanently with no possibility of alteration.

14. SUBJECT TERMS
Protecting Privacy, Terrorist Tracking, Inference Control, Access Control, Immutable Audit Trails

15. NUMBER OF PAGES
8

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Table of Contents

1.0 Introduction 1

1.1 Project Overview 1

2.0 Inference Control 1

3.0 Access Control 2

4.0 Immutable Audit Logs 3

5.0 Measurements of Effectiveness against a Synthetic Database 4

6.0 Bibliography 4

1.0 Introduction

This is the final report associated with Contract No. F30602-03-C-0037. With this report and the CDs delivered previously, we have completed the technical requirements of the contract.

This report is structured as follows, Section 1.1 provides an overview of the project, Section 2 describes our inference control work, Section 3 discusses our access control techniques, our method for generating immutable audit logs is in Section 4, the measurements we took of our tools are described in Section 5 and we conclude in Section 6 with references to all the research papers published under this contract.

1.1 Project Overview

Countering terrorism involves gathering information from a wide diversity of sources to discover key facts and relationships, develop models of hypotheses, and support human reasoning on likely futures and outcomes. Many of these data sources contain sensitive personal information, such as data on telephone calls, email, credit card usage, bank accounts, car rentals, housing, educational data, health-related data, drivers' licenses, airline and hotel reservations, visas, border crossing, attendance at events, and application for government programs. In tracking potential terrorists and attempting to discover their relationships and organization, is it necessary to focus on data about individuals. Yet it is this identification of data with individuals that makes the information sensitive.

The goal of this project was to allow authorized analysts to search these data for terrorist-related activity while providing a realistic degree of privacy protection for ordinary citizens who may be also represented in those databases. The proposed solution has the following elements:

Inference control to prevent unauthorized individuals from completing queries that would allow identification of ordinary citizens.

Access control to return sensitive identifying data to appropriately authorized users.

Immutable audit logs that ensure all data accesses are recorded immediately and permanently with no possibility of alteration.

2.0 Inference Control

In a multilevel database an inference problem exists if users are able to infer sensitive information from a sequence of attribute queries that each have a low security classification (i.e. are not sensitive). For example, any user may be able to query a database to retrieve a list of ship names and the ports at which they are docked. In addition, the knowledge of which ports are being used to load ships with weapons may have a low security classification. Yet, these two queries together constitute an *inference channel*, because if both are made then it's possible to infer exactly which ships are carrying weapons, and this may be sensitive.

Our inference control tool takes as input an annotated database schema that specifies identifying attributes, quasi-identifying attributes and functional dependencies between attributes. The inference control tool produces as output a file in XML format that lists all identifying attributes and inference channels. The code for our prototype inference control tool has been delivered to DARPA.

3.0 Access Control

When protecting against inferences, there is an inherent trade-off between the granularity of the access control and the query processing time. A common approach is to raise the security levels of specific objects in the database in order to prevent a user with low security clearance from completing enough queries to be able to make an undesired inference. By ensuring that at least one object in each inference channel requires high clearance, low-clearance users are prevented from making inferences. However, a user who only wants to query one particular object whose security level has been raised will be unable to do so without receiving additional authorization, even though the information they seek may be completely innocuous on its own. Hence, because access controls are predetermined, this approach may impede access to information unnecessarily.

Another approach to access control is to determine at query time whether a query can be safely answered. This can be done, for example, by maintaining user query histories. When a user makes a query it is checked against the user's query history and all known inference channels, before granting access to the results of the query. However, since query histories can be quite long, this approach can result in slow query processing.

We have developed a new approach that allows for fast query processing while enabling fine-grained access control, and thus, flexible information access. In our approach, access-enabling tokens are associated with objects in the database, and users are allocated keys that they use to generate the necessary tokens. Once a token is used to query an object the key it was derived from cannot be used to query any other object in the inference channel. This is implemented by deleting (or, revoking) the tokens generated with this key from other objects in the channel. Hence, query processing depends on the length of the channel rather than the ever-growing user query histories. In addition, because initially the same tokens are associated with each object, our approach allows for flexible information access. A user can access any objects in the inference channel provided doing so will not enable the user to make the undesired inference, even through collusion with other users.

Our approach to inference control is inspired by cryptographic revocation techniques for large groups of users. The motivating intuition behind the use of these techniques is that group dynamics play an essential role in ensuring access control: it is not enough to only consider the user requesting the object when deciding whether or not to grant access, instead all of the users of the database and all of the queries they've made, should somehow be taken into account. The difficulty comes in finding a way to do this without relying on the time-consuming processing of user query histories. As an example, one might imagine solving the problem by associating counters with objects in the channel, and cutting off access to the channel when the counters get too high. However, the counters reflect *x* queries by 1 user and 1 query by each of *x* users in the same way, and this doesn't sufficiently capture the access dynamics. By leveraging ideas from group communications we are able to provide some separation between these cases and an automated mechanism for updating the access controls that is far more likely to be affected by large-scale querying by a few users, than scattered queries by many users. More precisely, our schemes provide collusion resistance and a desirable new property that we call *crowd control*. Crowd control ensures that if a large number of users have queried all but one of the objects in the channel then no one will be able to query the remainder of the channel even if they have never queried the database before.

A research paper describing our access control techniques [2)] and the Java source code of our prototype access control mechanism for the privacy appliance have both been delivered to DARPA.

4.0 Immutable Audit Logs

System logs provide an invaluable view into the current and past state of almost any type of complex system. Most server software in existence today includes some logging mechanisms.

Secure versions of such logs, designed to defend against malicious tampering, allow the current state of the system to be audited even when that system has been under active attack by malicious insiders or outsiders. Correctly designed secure audit logging mechanisms can detect unauthorized past activity, even when the person performing that action goes to great lengths to cover their tracks. The existence of such logs can be used to enforce correct user behavior, by holding users accountable for their actions as recorded in the audit log. Such logs can be used in a wide variety of systems, from a control system that logs the commands a user issues, to a database system that logs the queries a user makes.

Typically, when an organization wishes to inspect past activity it will search the audit log for relevant information. For example, if a certain user was suspected of behaving improperly the organization might search for all actions performed by that particular user. If the organization wishes to see all actions of a certain type, it might search for all log entries that match a given keyword. For an audit log to be useful in practice, it is critical that it be efficiently searchable for keywords of interest.

At the same time, the contents of an audit log can be considered to be sensitive information. For instance, knowing what actions are made by a certain user could violate that individual’s privacy. If the log contains information about not only what query was made, but what results were returned, access to the audit log would imply effective access to the database, circumventing database access controls. The organization that owns the system being logged might consider the information the log holds to be valuable and not wish to share it with others, while for robustness’ sake, the organization may want to store backup copies of the audit log information at sites it may not completely control. In general, this means that the contents of the audit log must be encrypted. However, this makes it extremely difficult to search. Using traditional techniques, searching the log would require decrypting every record. This approach has several disadvantages. First, it requires decrypting all of the log data, regardless of what information one is looking for; this opens opportunities for unintended access to log records other than the ones relevant to the current investigation. Second, it requires the entity with the decryption key to interactively process all the log data, which can be quite large. In many applications, one would like to entrust the ability to decrypt audit logs to an entity or system with high levels of trust and assurance; requiring that system to also be able to process large quantities of log data in an on-line fashion limits one's choice of trusted parties. It would be preferable to be able to selectively delegate the ability to search the log to parties with the means to process the data.

The key challenge to building a successful, secure audit logging system is to simultaneously protect the integrity of the audit log, control access to contents, and maintain its usefulness by making it searchable.

We have developed a method for generating encrypted audit logs that allows a designated trusted party, the audit escrow agent, to construct keyword search capabilities, which allow (less trusted) investigators in possession of such capabilities to search for and decrypt entries matching a given keyword. The escrow agent can distribute a capability to an investigator if he deems it appropriate. Since we expect keyword search capabilities to be distributed rather infrequently, the escrow agent can be made to be very secure from attack.

The research paper [3)] explaining our audit trail technique and the corresponding C++ source code, have been delivered to DARPA.

5.0 Measurements of Effectiveness against a Synthetic Database

We measured the effectiveness and overhead of our access control tool and immutable audit trail against a synthetic database. We issued a set of 24 queries to the database, coming from 5 fictitious users of the database. The queries, the query responses and the measurements of performance overhead and identity inference have been delivered to DARPA.

6.0 Bibliography

1) P. Golle, J. Staddon, B. Waters. Secure conjunctive search over encrypted data. *ACNS 2004.*
2) J. Staddon. Dynamic inference control. *8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*.
3) Brent Waters, Dirk Balfanz, Glenn Durfee, and Diana Smetters. Building an encrypted and searchable audit log. *Proceedings of the 2004 Network and Distributed Systems Security Symposium (NDSS'04)*. The Internet Society. San Diego, CA. February 2004.
4) D. Woodruff and J. Staddon. Private inference control. *ACM CCS 2004.*