Privacy management system using social networking

Song, Ronggong; Korba, Larry; Yee, Georg

Publisher's version / Version de l'éditeur:

The 2007 IEEE International Conference on Systems, Men and Cybernetics (SMC 2007), 2007

National Research Council Canada
Conseil national de recherches Canada

Institute for Information Technology
Institut de technologie de l'information

NRC·CNRC

*Privacy Management System Using Social Networking **

Song, R., Korba, L., and Yee, G.
2007

* Proceedings of the 2007 IEEE International Conference on Systems, Men and Cybernetics (SMC 2007). October 7-10, 2007. Montreal, Canada. NRC 49368.

Canada

Privacy Management System Using Social Networking

Ronggong Song, *Senior Member, IEEE,* Larry Korba, and George Yee, *Senior Member, IEEE*

***Abstract*—The worldwide growth of e-services has brought to the forefront the importance of private data management for an organization. However, the advancement of computer and networking technologies has made privacy management very challenging. In this paper, we expose the limitations of existing privacy management systems, and present a privacy management system that exploits social network analysis, which can automatically discover the privacy-related workflow models, and support automated privacy management within an organization.**

I. Introduction

PRIVACY management and violation are becoming serious issues with the worldwide growth of e-services (e.g. e-government, e-health, e-learning, and e-commerce, etc.) and the advanced computer and networking technologies. The growth of e-services and the inexpensive large capacity storage technologies allow an organization to very easily collect and store increasingly larger amounts of private data and related events in digital form from growing numbers of clients. In addition, organizations are deploying more and more mobile computers such as laptop and handheld computers due to their convenience and decreasing costs. Advances in computer and networking technologies enable staff to process the private data everywhere for delivering services and products, which can lead to distributed storage of the private data. The processing and storage of the private data everywhere makes privacy management very difficult for organizations.

Privacy management technologies have been under research for many years. However, most research focuses on protection of the private data stored in a particular location such as a centralized database where it can be monitored, for example, IBM's Enterprise Privacy Architecture [1] (EPA) and Camenisch et al's Privacy and Identity Management for Europe [2] (PRIME), etc. EPA is intended to help organizations minimize the risks of inadvertent privacy disclosure by showing them where personally identifiable information (PII) is stored in their enterprise and how to effectively manage it. In the technical architecture, EPA uses the privacy enforcement system to control disclosure of the private data. However, except for monitoring the activities in the indicated location where the private data is stored, based on the privacy agreement framework, the EPA technical architecture does not monitor privacy-related activities anywhere else in the whole organization to ensure enforcement of privacy policies and to prevent improperly processing by employees. For instance, an employee can save the private data in a local storage device during service delivery and misuse it later (e.g. copy it onto a USB memory, device or send it out by e-mail, etc.). PRIME (Privacy Identity Management for Europe Project) implements a technical framework for processing personal data that is meant to complement the EU's legal framework for the processing of personal data. It is investigating the idea of giving the user control over his private information for transactions with a service provider, and provides identity management on both user-side and service-side. The user side can control the amount of private information released to different service identities, and the service side protects the private data by access control and authorization. PRIME uses an obligation manager to monitor the activities on private data stored in the database, but it does not provide the techniques for comprehensive insider threat detection on private data stored anywhere within the enterprise. For instance, it does not prevent the activities of internal employees that would cause privacy violations (e.g. print the private data saved in the local storage during service delivery). Other privacy management systems are only applied to special applications such as P3P [3] and APPEL [4] for web-based applications. In addition, they do not satisfy many aspects of privacy legislation. For instance, they do not specify how the collected private data are stored and what security mechanisms are used for safeguards. Since privacy management is becoming an enterprise-wide issue a key starting point is discovering where the private data is, across the organization, and understanding how it is used everywhere in the organization. Obviously, existing technologies have limitations in this regard.

In order to manage the private data stored in all locations throughout the enterprise, we are researching and developing a privacy management system that makes use of techniques for discovering how and who deals with private data. Our system uses private data discovery, privacy-related events detection, and social networking analysis technologies to support enterprise privacy management. The system is comprised of several different agents installed in every computer throughout the enterprise in a distributed fashion. The agents can automatically discover the private data stored in a local computer and detect the activities related to private

Manuscript received April 15, 2007.

Ronggong Song is with the National Research Council of Canada, Ottawa, Ontario K1A 0R6, Canada (corresponding author to provide phone: 613-990-6869; fax: 613-952-7151; e-mail: ronggong.song@ nrc-cnrc.gc.ca).

Larry Korba with the National Research Council of Canada, Ottawa, Ontario K1A 0R6, Canada (e-mail:larry.korba@ nrc-cnrc.gc.ca).

George Yee with the National Research Council of Canada, Ottawa, Ontario K1A 0R6, Canada (e-mail: george.yee@nrc-cnrc.gc.ca).

NRC-49368

data handling using privacy data mining and privacy-related activity detection technologies. Using a peer-to-peer communication model, the agents communicate with one another through a local interface agent, which discovers local privacy-related workflow using social networking analysis technology to support privacy analysis and assessment. In addition, an enterprise privacy-related workflow is also produced by a data controller agent by using social networking analysis technology in order to understand how private data is being used throughout the enterprise and supporting the overall privacy violation analysis and enforcement of privacy polices.

The rest of this paper is organized as follows. Section 2 presents our proposed privacy management system that uses social networking, including the system architecture design and important techniques used in the system. Section 3 shows the system prototype implementation and results. Section 4 gives our conclusions.

II. Privacy Management System Using Social Networking

In our system, we combine several technologies that allow organizations to manage their private data and understand how it is being used throughout the enterprise. The technologies cover the management of privacy policies, data mining to discover the private data and detect the privacy-related events, and social network analysis to discover the privacy-related workflow and analyze the private data usage.

A. System Architecture

The system uses a peer-to-peer networking architecture and is comprised of several agents: data controller agent, local interface agent, privacy monitoring agent (PMA), privacy assessment agent (PAA), and privacy enforcement agent (PEA). The local interface agent, privacy monitoring agent, privacy assessment agent, and privacy enforcement agent are installed on every computer within the enterprise. The privacy monitoring agent can automatically discover the private data stored in the local computer and detect the privacy-related events in that computer by the privacy data mining and privacy-related activity detection technologies. A PMA can communicate with other agents via the local interface agent, which produces local privacy-related workflow correlating activities on related private data to support the privacy assessment agent and privacy enforcement agent. The data controller agent manages the privacy policies and related private data, and produces enterprise privacy-related workflows based on social networking analysis models applied to all privacy-related events collected from local privacy monitoring agents. Data visualization techniques are applied to the privacy-related workflows to improve understanding of how the private data is being used within the organization and highlight problem areas. In addition, the data controller agent controls other agents' behaviors for privacy compliance analysis and enforcement. An organization can install one or more data controller agent in the system depending on the size of the organization. Figure 1 depicts the general architecture of the system.

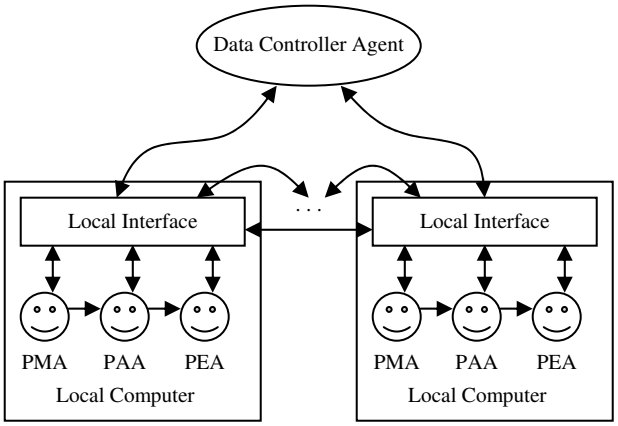

Fig. 1. The general system architecture.

B. Important Techniques

The key techniques used in the system include computer activity monitoring, private data scanning, private data and events detection, and social network analysis. Figure 2 depicts the abstract model of the key techniques used in the system.

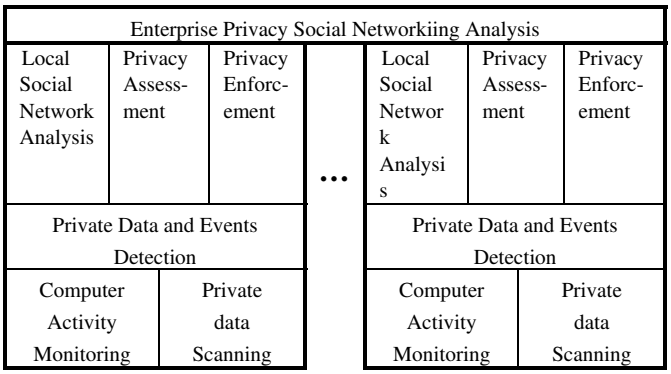

Fig. 2. The abstract model of key techniques used in the system.

Computer activity monitoring involves monitoring raw system calls to capture all activities in real time on the computer, including windows operation, file operation, keystrokes, and packet traffic. However, the privacy monitoring agent only reports the privacy-related events by filtering the monitored events for only those related to private data.

Private data scanning is used to search the private data stored in the file system of the local computer using private data detection patterns, mainly using regular expression matching. So far, we have developed identity-related private data detection patterns to detect data such as Social Insurance Numbers (SINs), contact-related private data detection patterns such as postal addresses, email addresses, and

telephone numbers, and financial-related private data detection patterns to detect financial data such as credit card accounts and bank accounts.

The private data and events detection filters the computer activities, which are monitored by the computer activity monitoring engine, with the private data detection patterns and private data mining techniques.

The social network analysis applies the data mining techniques and social network models to the collected private data and related events to build private data workflow for understanding how the private data is being used throughout the organization and support the privacy assessment and enforcement.

C. *Agent Design*

Based on the above system architecture and key techniques used in the system, we designed the functionality models of each agent as follows.

Data Controller Agent

The Data Controller Agent is designed to communicate with the data subject and private data administrator, manage the private data and privacy polices, produce the enterprise privacy social networking analysis models, and control other agents. It contains the following major components: data controller user interface, registration and access control, privacy policy management, enterprise private data workflow visualizations, privacy compliance agents controller, protected private data, and enterprise privacy social networking analysis models. Figure 3 depicts the virtual functionality model of the data controller agent. The detailed information about these functionality components is described as follows.

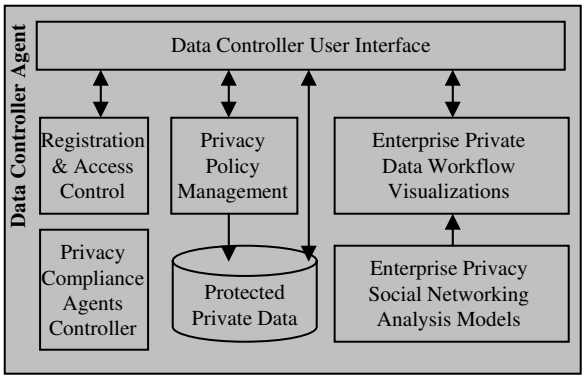

Fig. 3. The virtual functionality model of data controller agent.

- Data Controller User Interface: This is a GUI interface by which the data subject and enterprise privacy administrator can access the data controller server to manage the privacy policy and private data, and review private data processing performance via enterprise private data workflow analysis and visualizations.
- Registration & Access Control: This provides registration and authentication services for the data subject and enterprise privacy administrator to securely access the data controller server. It supports the interface for creating a secure channel between the remote user and data controller server.
- Privacy Policy Management: This provides the privacy policy management service for the data subject and enterprise privacy administrator to manage their privacy policies. Privacy policy negotiation technologies [5] may be applied to allow the data subject to negotiate his privacy policy with the data controller.
- Protected Private Data: This provides the private data management service for storing the private data into the database of the data controller's server. The private data is protected and controlled by the protection mechanisms and privacy policy.
- Privacy Compliance Agents Controller: This provides the agent control service for the data controller to control the local interface agent, privacy monitoring agent, privacy assessment agent, and privacy enforcement agent that are running on the local computer.
- Enterprise Privacy Social Networking Analysis Models: This provides the social networking analysis mechanisms to determine the correlation among the private data processing events within the enterprise and build the enterprise private data workflow. It manipulates the models based on a set of methods and provides them to the private data workflow analysis and visualization for privacy compliance assessment.
- Enterprise Private Data Workflow Visualizations: This provides the visualization tools for displaying the social networking analysis results on the private data processing events collected by the local privacy monitoring agent.

Local Interface Agent, Privacy Monitoring Agent, Privacy Assessment Agent, Privacy Enforcement Agent

These agents are installed in all computers located throughout the organization. Figure 4 depicts the virtual functionality models of each agent.

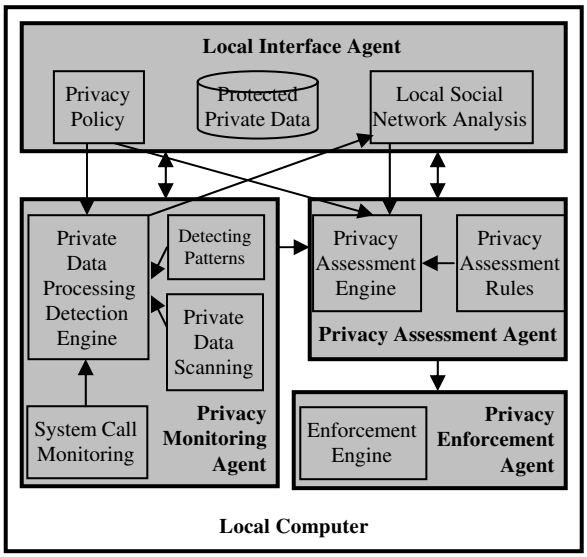

Fig. 4. The virtual functionality models of the agent on the local computer.

The Local Interface Agent is designed to communicate with the data controller agent, requesting the privacy policy and private data used in the local computer, and forwarding the collected local privacy events and local social networking analysis results to the data controller agent. It also communicates with other local interface agents to request the related privacy events for building local social networking analysis models.

The Privacy Monitoring Agent is designed to detect the privacy-related activities on the local computer and report the private data processing events to the data controller. It signals the Privacy Assessment Agent for privacy compliance evaluation on the detection of a private data processing event. The PMA also contains the following major components: system call monitoring, private data processing detection engine, private data processing detection patterns, and private data scanning. Details about these components are disclosed as follows.

- System Call Monitoring: This provides the computer activity monitoring service for the PMA, monitoring all activities on the local machine in real time, including system operations, application and file operations, keystrokes, and network traffic. As a demonstration under Microsoft Windows, we developed a) FMon for monitoring file operations such as file opening, closing, editing, deleting, and saving, b) WMon for monitoring windows operations such as window opening, closing, activating, title changing, file system navigation recording, and Internet navigation recording, c) KeyStrokeMon for capturing all keystrokes, and d) PacketMon for capturing all Internet traffic packets.
- Private Data Processing Detection Engine: This provides the privacy detection service for the PMA to check whether an activity event is related to private data. The detection engine is controlled by the detection patterns and privacy policy. The detected private data processing events are immediately forwarded to the local interface agent for building local social networking analysis models. Meanwhile, a control signal, together with the detected events, is sent to the Privacy Assessment Agent for near real-time privacy violation evaluation.
- Private Data Processing Detection Patterns: This is a set of private data processing detection patterns to control the private data processing detection engine, for instance, the telephone number detection pattern, credit card number detection pattern, and others.
- Private Data Scanning: This provides the privacy compliance scanning service for the PMA, by scanning the file system determining whether the system protects private data properly (e.g. whether the private data has potential for outside leakage or has already leaked out). It uses the privacy compliance scanning engine to scan a variety of file types such as PDF, DOC, PPT, RTF, TXT, and others. The scanning engine is controlled by the private data detection patterns.

The Privacy Assessment Agent is designed to evaluate whether the detected private data processing events violate the privacy policy or privacy legislation principles. The assessment is based on the privacy compliance assessment rules, privacy policy, privacy legislation principles, and privacy social networking analysis results. The assessment results will be sent back to the data controller for further analysis. This agent will either signal the Privacy Enforcement Agent to stop the processing or alert the data processor once a privacy violation event is detected through the assessment. The PAA also contains the following major components: privacy assessment engine, privacy principles service, and privacy assessment rules. Detailed information about these components is described as follows.

- Privacy Assessment Engine: This provides the privacy compliance evaluation service for the agent to evaluate whether the detected private data processing events received from the detection engine comply with the privacy policy and privacy principles. The privacy assessment engine is controlled by the private data processing events detection engine, privacy policy, privacy principles, and privacy assessment rules. It signals the PEA to stop the processing or alert the data processor once the assessment result shows that the processing is a privacy non-compliance event.
- Privacy Principles Service: This provides the privacy principle management service for the agent to get the privacy principles based on the privacy legislation (e.g. EU Directive [6], Canada PIPEDA [7]) and further use it to control the privacy assessment engine.
- Privacy Assessment Rules: This is a set of privacy compliance assessment rules to control the privacy assessment engine. For instance, these rules may infer the unauthorized private data access and editing events, identifying the potential leakage of the private data, and classifying the privacy non-compliance processing events at different privacy violation levels, and others.

The PEA is designed to stop the private data processing or alert the data processor that activities have been detected that violate the privacy policy. The privacy assessment engine of the PAA sends a control message with the detected event and related privacy assessment result to the privacy enforcement engine once the privacy non-compliance event has been detected. The privacy enforcement engine takes over the system operation and informs the data processor about privacy violation processing according to the non-compliance levels of the detected events.

III. Implementation And Results

Our system prototype is implemented in C++ and Java, and based on the Microsoft Windows XP Operating System. It uses the Java Agent Development Environment (JADE) [8] to provide a software development environment, and SPKI/SDSI [9] to provide the distributed security infrastructure for secure communication between each local interface agent. The data controller agent runs in the main container of the data controller's machine. Other agents run in

the sub-containers of the local computers. In the following, we only introduce several key technologies used for the prototype development, since many components of the architecture are still undergoing research or are being further developed.

Computer Activity Monitoring

As described above, activity monitoring monitors all activities on the computer in real-time including windows operations, file operations, keystrokes, and packet traffic in order to get enough information related to private data events analysis. The following Java classes are developed for real-time activity monitoring under Microsoft Windows XP:

- FMon: File monitoring class using JNIWrapper [10], which is a software development kit working with native code libraries within Java,, together with the file system watcher feature of WinPack [11] to monitor the file activities such as file creation, rename, edit, and deletion.
- WMon: Windows monitoring class using system–wide Windows Call mechanisms to monitor windows events such as window opening, closing, activation, and others.
- KeyStrokeMon: Keystroke monitoring class using the keyboard hook feature of WinPack combined with JNIWrapper to capture the data subject's keystrokes.
- PacketMon: Packet monitoring class using Jpcap [12] to capture the network traffic flowing through the computer.

All the above classes are implemented with Java, and controlled by an activity monitoring interface class – ActivityMonitor. Figure 5 depicts the class diagram for the activity monitoring. For the implementation of activity monitoring, the big issue is scalability.

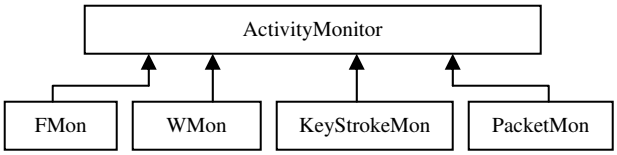

Fig. 5. ActivityMonitor Class and Sub-Classes

Private Data Events Detection

In order to make the system more efficient, accurate, scalable, and to provide enough information required for privacy compliance analysis and forensics, we developed the following data structure for the privacy processing events. Figure 6 depicts this data structure.

Privacy Event, the privacy processing event entity, is an aggregation of eight other entities:

- Data Subject: The person or entity to whom the private data belongs. This may also be a reference to the data subject (e.g. number),
- Data Processor: The processor that worked on the private data to create a processing event,
- Personal Data: The private data that was involved in this event. This part of the record contains two fields: Type and Value if the private data is classified [13],
- Application: The application within which the data processor processed the private data. It contains two fields: Value (e.g. IE, Outlook) and Context. The Context value is to locate the detailed information describing the private data storage (e.g. file name, URL, e-mail subject),
- Operation: The processing that the data processor performed on the private data in a processing event. It contains two fields: Value (e.g. read, modify, copy, send) and Context to gather contextual information related to the operation's target (e.g. e-mail address, FTP server),
- Computer: The computer on which the data processor performed the private data processing. This may contain the computer's IP address, MAC address, or machine name,
- Time: The time of the data processor operation,
- Extensions: These are fields for further additions.

These data can be collected efficiently in our prototype with FMon, WMon, KeyStrokeMon, and PacketMon. One private data event record may be composed of data from multiple monitoring engines after applying data correlation techniques.

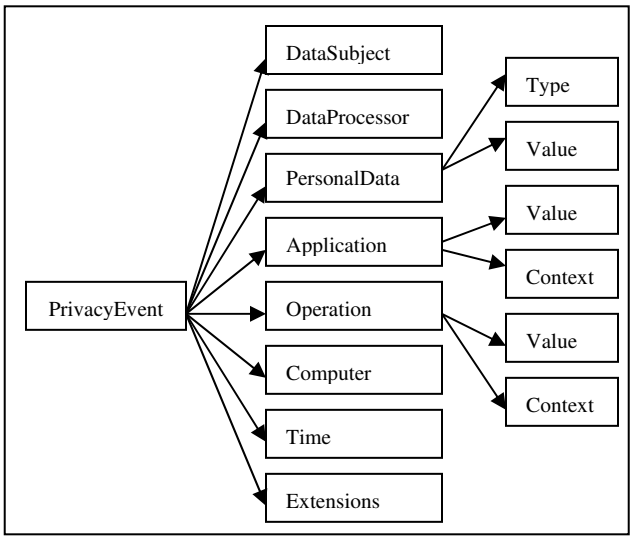

Fig. 6. Privacy processing event data structure.

Privacy related events and data detection technologies are used to filter the monitored activities for privacy-related events collection. As we mentioned, the key technologies for this purpose include the private data mining and privacy-related events detection patterns. For the prototype, we have implemented some detection patterns to identity, financial, and contact data such as social insurance number (SIN), credit card number, bank account number, phone number, postal address, and e-mail address. Figure 7 shows an example of a private data processing detection report from the prototype. However, handling the wide possible range of patterns related to privacy processing events needs further development. In addition, detection accuracy and efficiency are very important factors for the design of detection patterns and their implementation.

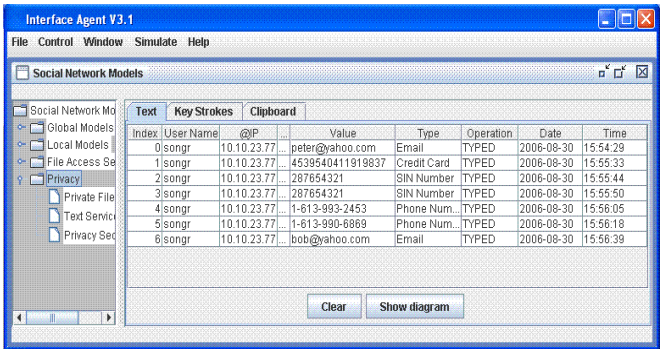

Fig. 7. Private data event detection report.

Social Network Analysis and Visualization

Social network analysis is used to correlate and cluster the detected privacy events, infer the relationships among them, and build social network models based on different views and conditions by applying correlation and clustering rules. The current prototype can produce visualizations of two different models: a relative cluster model, and a time sequence model. Figure 8 shows a cluster model with 3 data processors sharing private data artifacts (targets of the arrows), showing relationships among the data processors based on how they use and share private data. Figure 9 shows a time sequence model with correlation rules, which builds a workflow (sequence of blue dots) to display how the private data has been used by the data processors based on the time sequence.

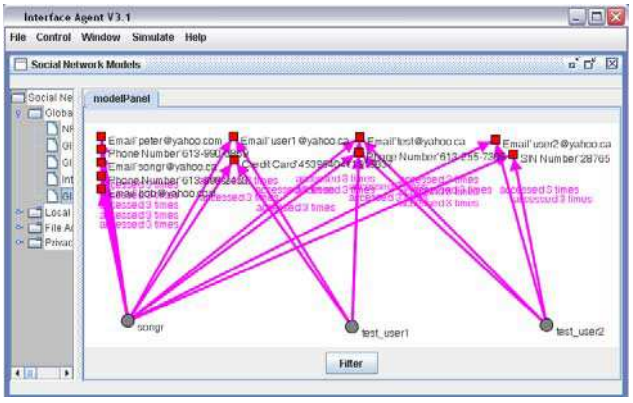

Fig. 8. A cluster model of the detected privacy processing events.

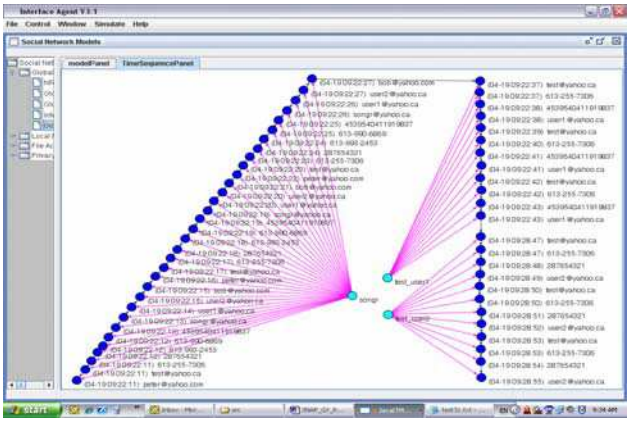

Fig. 9. A time sequence model of the privacy processing events.

IV. Conclusion

With the proliferation of e-services, privacy violation is becoming a serious issue. In order to provide better protection for private data, force private data processing to comply with privacy policy and legislation throughout the enterprise, we propose SNAP, a privacy management system that uses social networking. In this paper, we have provided a description of the SNAP architecture and discussed its prototype implementation. However, many technologies related to the architecture such as private event detection patterns, privacy compliance assessment, and so on, need further research.

Acknowledgment

The authors acknowledge the support of the other SNAP project team members: Andrew Patrick, Scott Buffett, Yunli Wang, George Forester, Marc-Alain Mallet, and Sharon Wahl, and the development contributions of Rougu Lou.

References

[1] IBM EPA. Enterprise Privacy Architecture (EPA). From http://www.zurich.ibm.com/pri/projects/epa.html Retrieved April 6, 2007.

[2] J. Camenisch, A. Shelat, D. Sommer, S. Fischer-Hubner, M. Hansen, H. Krasemann, G. Lacoste, R. Leenes, and J. Tseng. Privacy and Identity Management for Everyone. Proceedings of the 2005 Workshop on Digital Identity Management, pp. 20-27, Fairfax, VA, USA, 2005

[3] L. Cranor, B. Dobbs, S. Egelman, G. Hogben, J. Humphrey, M. Langheinrith, M. Marchiori, M. Presler-Marshall, J. Reagle, M.Schunter, D. A. Stampley, and R. Wenning. The Platform for Privacy Preferences 1.1 (P3P 1.1) Specification. From http://www.w3.org/TR/P3P11/ Retrieved April 6, 2007 .

[4] L. Cranor, M. Langheinrith, and M. Marchiori. A P3P Preference Exchange Language 1.0 (APPEL 1.0). From http://www.w3.org/TR/P3P-preferences/ Retrieved April 6, 2007.

[5] G. Yee and L. Korba. Negotiated Security Policies for E-Services and Web Services. Proceedings of the 2005 IEEE International Conference on Web Services (ICWS 2005). Orlando, Florida, USA. July 11-15, 2005. NRC 47449.

[6] EU Directive 95/46/EC. EU Official Journal of the European Communities, No L. 281, 23/11/1995 p/0031—0050.

[7] Office of Privacy Commissioner of Canada. The Personal Information Protection and Electronic Documents Act. From http://www.privcom.gc.ca/legislation/index_e.asp Retrieved April 6, 2007.

[8] JADE - Java Agent Development Environment. From http://sharon.cselt.it/projects/jade Retrieved April 6, 2007.

[9] SDSI – A Simple Distributed Security Infrastructure. From http://theory.lcs.mit.edu/~cis/sdsi.html Retrieved April 6, 2007.

[10] JNIWrapper, TeamDev Led. From http://www.teamdev.com/jniwrapper/index.jsf. Retrieved April 6, 2007.

[11] WinPack, TeamDev Led. From http://www.jniwrapper.com/pages/winpack/features Retrieved April 6, 2007.

[12] Jpcap – Java Package for Packet Capture. From http://netresearch.ics.uci.edu/kfujii/jpcap/doc/index.html Retrieved April 6, 2007.

[13] R. Song, L. Korba, and G. Yee. Privacy Rights Management for Privacy Compliance Systems. Proceedings of the IEEE 21st International Conference on Advanced Information Networking and Applications (AINA-07) Symposia – The Third IEEE International Symposium on Security Networks and Distributed Systems (SSNDS 07), Niagara Falls, Canada, May 21-23, 2007.