The objective of this project was to design techniques that preserve confidentiality and integrity of information in computer systems while providing dynamic trust-based updates so that information is more readily available. We have developed various models including for managing data in the web of trust, mapping objects to various trust zones, assigning trust values to objects based on their component structures and subjects' evaluations, identification of corrupted objects in the system, determining vulnerability of subjects by deceptive information in an information flow network, data authentication and provenance, storage and management of provenance metadata, restricting inferences of sensitive data from non-sensitive data, and knowledge extraction and analysis for insider threat mitigation. This work resulted in three Ph.D. dissertations, 11 M.S. thesis, and 27 research papers with few more in press for publication.

Object trust, management of object trust, trust zone mapping, information flow, information dissemination model, data authentication, inference problem, insider threat mitigation.
Final Report
for
DEPSCoR 2004 Project

Submitted to
The Air Force Office of Scientific Research

by
The University of Arkansas

Proposal Title: Trust-based Hierarchical Role Enhanced Policy for Adaptive Availability of Confidential Information

Grant Number: FA9550-04-1-0429

Principal Investigator: Dr. Brajendra Panda

Program Manager: Dr. Robert L. Herklotz

Major Accomplishments: The following key research activities have been accomplished as a result of this project.

- Model to evaluate effect of deceptive data in a web of trust.
- Computer trustworthiness of objects based on trustworthiness of their components and recommendations provided by different subjects.
- An information assurance model to map objects received from external sources to appropriate trust zones.
- A model of interpersonal trust to allow varying levels of information flow between peers.
- A policy-oriented trust model that indicates which features of external objects are favorable and which features are undesirable.
- Development of a formal data structure to show how a given piece of information was formed.
- Algorithms to compare the component structure similarity/dissimilarity between two object versions.
- A method to identify all corrupt objects in the system when a user maliciously cause damage to one or more data objects.
- Information dissemination model based on both the information flow network and the web of trust to determine effect of malicious activities by untrustworthy users.
A Trust-based Two-way Information Dissemination model to study the effect when people voluntarily push information through the network and when people send information only when they are requested.

Principles and methodology for management of object trust to help users design trusted computer systems.

Data authentication and storage methodology

Three different models to solve the aggregation inference problem in order to restrict unauthorized user from deducing sensitive data from non-sensitive data.

An insider threat analysis model based on each insider’s knowledge

Dr. Hexmoor, the co-PI of this project, left the University of Arkansas in August 2006. So, we had to re-organize his part of the research and that resulted in a delay in finishing the project.

Executive Summary

For development of trust-based policies, we started with the web of trust, which is a building block for the semantic web and e-commerce applications. Our idea was to study how deceptive data spreads in a web of trust model rather than focusing on traditional research on trust rating of subjects and objects. Any deceptive data, when sent by a highly trusted user, can not only affect people who directly trust the sender, it would have cascading effect on a number of other people in the network. Our research focused on evaluation of the effect of deceptive data and the extent to which people in the trusted network may be affected by such data. The model we have developed illustrates how the web of trust and information flow network can be used in conjunction to assess the detrimental effect of deceptive data. We also have used the concept of community and personal social circle properties of web of trust to illustrate how the structural analysis of web can help evaluate the result of spreading deceptive data. By identifying the group of people that are affected by the data, appropriate strategy to recover from the effect can be developed.

As our next task, we studied trustworthiness of objects in virtual organizations. One of the difficulties in evaluating their trustworthiness is the lack of sufficient information to see how the object was formed and to what level its components should be trusted. Users need to be provided with information on the structure of a compound object in order to evaluate the trust level of that object. The model we developed introduces a technique for trust management using labels associated with each object within the domain of a virtual organization. Each label supplies certain information regarding the originality of the associated object. Thus, partial trust (also called component trust) can be integrated to evaluate the composite trust of compound objects. Different subjects may view the same object with different trust values since they trust the object’s components to different degrees. Our model uses recommendations provided by various subjects to compute the trustworthiness of an object for a given subject.

In a loosely-coupled system various objects may be imported from different sources and the integrity levels of these objects can vary widely. Like downloaded information from the World Wide Web, these imported objects should be carefully organized and disseminated to different trust zones, which meet the security requirements of different groups of internal
Assigning an object to a trust zone is called trust zone mapping, which is essentially a form of information clustering and is designed to guide internal applications when they use objects from different zones. We have developed a model for information assurance by mapping external objects to appropriate trust zones. This mapping serves two purposes: limit access rights of external executable programs to internal resources and guide internal applications to use trusted external information. We have defined two powerful threshold selection operators to check and verify if an external object satisfies the trust-based security conditions as specified by each trust zone. Primary and secondary trust values for an object are calculated by our method.

We have developed a model of interpersonal trust. The technique uses the notion of boundary spanner from organizational theory to model a central point of trust from an organization projected outward. We have introduced a method for augmenting online communities with security. The method extends the “friend of a friend” protocol enabling formation of secure, peer-to-peer community initiation and trust policies that allow varying levels of information flow while protecting integrity of information exchanged. We have developed methods for detecting unusual, “suspicious”, patterns of trust in a P2P network.

We have also devised a policy-oriented trust-based decision model for subjects to select reliable and secure information in an open system. Achieving security and quality of service is important for open systems where there is no single authority and where traditional security models do not work effectively. Our model allows a user to specify what features of external information it can’t accept and what features are favorable to it. Based on this model, an example of a policy specification has been defined. Selector, a high-level policy language, has been developed to express the user-defined policy specification that allows automatic evaluation of the trustworthiness of available object versions of a given object and select one that meets the user’s requirements for information quality and security. The work also introduced object trustworthy calculations, which are important for users to make trust decisions. Compared with other decision-making approaches, our trust selection model is easy to understand and can be applied in computing systems. The model allows users to specify their customized policies to address their concerns for information integrity.

Given a specific version of an object, we have derived a method that allows an evaluator to use first-hand information in evaluating its trustworthiness and the trust value calculated is called the primary trust value of the object. In case the user thinks that it is difficult to derive the primary trust value, (s)he may compute the secondary trust value by first calculating the primary trust value of the corresponding compound object version. This second method is much more efficient than the first one as it does not require recursion. Furthermore, based on the component-based approach, we have designed two heuristic methods, which can be used to estimate the trustworthiness of an object version. They are not to replace the general object trust method but serve as complementary approaches to the computations of object version trust.

When some information is derived from various data items gathered from multiple sources, it is possible that no data value may satisfy an evaluator’s requirement with regard to information quality, if they are evaluated separately. In order to verify information legitimacy and accuracy, we have developed techniques that study and compare intrinsic features of the
available information, i.e., consider the object values provided and the way the information has been computed. Our method is based on the “multiple-proof” logic. The developed technique is very much valuable in the environment where participants of a virtual organization have different levels of expertise and information processing culture, thus making it difficult to evaluate the quality of information they provide. By using a formal data structure to represent how a piece of information was obtained from various components, our model computes the trustworthiness of the information.

In case an untrustworthy user accesses data objects, there is a chance that (s)he may corrupt data objects, intentionally or unintentionally. We have developed a model, which uses data dependency relationships in user tasks or transactions to monitor unauthorized actions by users. Dependencies are determined by using the read, pre-write, and post-write sets of data items, which are generated by the static and dynamic semantic analyzers. User applications or database logs can be checked to construct these sets. By finding the data dependencies among transactions, we identify anomalies hidden at the user task level. A Petri-Net based implementation concept has been designed to check these kinds of data correlations at user task level as opposed to the transaction level. Once a malicious activity is carried out, all data objects modified by the user are considered corrupted. When an object is modified based on the values of a corrupt object, the former also gets damaged. To identify such objects in a distributed database system, we have developed several approaches that use log files to identify data relationships and accurately determine the list of all damaged objects.

In order to analyze deceptive information flow, especially the way deceptive information flows among the subjects in the web of trust, we have designed an information dissemination model. It determines the prerequisite for information dissemination based on both the information flow network and the web of trust. We have also developed the technique for evaluating the spread of deceptive data with polynomial algorithms. By conducting experiments we have come up with some interesting characteristics of the web of trust that affect the dissemination of deceptive data. The results reveal that although the increased outdegree of the originator of deceptive data can contribute to the augment of the number of affected subjects, the average outdegree of the nodes in the web of trust plays a major role in determining such number. Furthermore, we also learned that the number of affected subjects does not increase linearly with the increase of the number of nodes in the web of trust. These important results help in determination of the range of spread of untrustworthy information among users of an information system.

In addition to the above, we have developed a Trust-based Two-way Information Dissemination model having two sub-models, which we call the Push Model and the Pull Model. While in the former, people voluntarily send information to others, the latter is restricted to information flow only when a user requests such information. An interesting feature of this model is that it illustrates how trust relationships in the web of trust affect the dissemination of the deceptive information. We realized that when the originator of the information is identified by the recipient, the trust ranking that the receiver gives to the message is only dependent on the web of trust regardless of particular information flow path that actually carries the message from the originator to the final receiver. However, if the originator cannot be identified, the trust
rating of a message given by a receiver is dependent on the trust rating of the sender (who may not be the originator of the information) as assigned by the receiver.

We have designed a framework of object trust management and produced two object trust principles in an open system such as a virtual organization. The object trust principles specify reasoning and guidelines for information assessment. Our method allows users to select the information with the required level of quality and security features. Studying the trustworthiness of external information is challenging since it requires the evaluator to possess solid domain knowledge about that information. Our object trust principles provide formal methodologies and strategies to design autonomic or semi-autonomic and trusted computing systems and applications to assess whether a given object has the required level of quality and security features indicating their trustworthiness.

With a view to providing trustworthy data to users, we have devised a data authentication method along with a provenance storage mechanism. Various methods have been proposed to manage the provenance information of a data item as well as a broad investigation of the factors that might affect a user’s decision as to what approach needs to be adopted based on the user’s preferences and needs. We also have developed an efficient methodology to store the provenance information of data items that greatly facilitate the user’s conceptual perception of the storage methodology. The authentication technique makes users aware of the level of trust they can associate with any piece of data in question. Our model includes a component to compute the data reliability rate, which is then forwarded to users. We also have developed a methodology to store the provenance information of data items that is based on a tree structure where the root of the tree is a particular abstract data item, the children of which contain the data items whose provenance information might be needed. Our developed technique offers the following three advantages. First, the tree can have at most three levels: the top level is the root, the second level consists of the instances of the abstract data item, and the third level is made up of the data items that are part of their parent’s provenance information. Hence, searching the tree is relatively fast since it is made up of at most 3 levels. Secondly, the complexity of the structure is very reasonable since it does not involve much pointer manipulation. Finally, the structure is simple enough for the users to understand and manage the storage structure.

Inference control is a primary issue in databases that contain sensitive data. The key mode for inference in many databases is aggregation. We have come up with three models that are built upon each other to solve the aggregation inference problem. The first model is the base model, which does inference control by maintaining an Inference Dispersion (Δ) for each user. A threshold is set on the value of Δ, and users whose Δ value exceeds the threshold are not sent any more data items. However, the maintenance of only a single Δ value to a user leads to less accessibility to the users. So, to solve this problem, we introduced the second model that separates the Δ value for each aggregation graph associated with each user. The presence of a single inference interpreter in these two models lead to issues such as slower query processing, single point of failure, and other problems associated with stand alone systems. These problems were solved using the third model based on distributed processing. The advantages of these models are that they are simple to implement and are domain independent.
In order to deal with insider threat problem, we have devised a technique that enables an organization control its insiders. Our method makes use of the fact that the primary difference between an insider and an outsider is the knowledge the former has gained by working for the organization. Understanding the knowledge a user acquires by accessing a data object and using this information to control the user’s future activities are the basis of our model. We use an ontological approach to extract knowledge units from each data object. All such knowledge units and their relationships are represented in graphs called Knowledge Graphs. Similarly, relationships among various data objects in the system are captured using a Dependency Graph. We have developed an algorithm for insider threat prevention that uses the above two types of graphs to ensure that an insider accesses only objects that are related to his/her domain and assigned tasks. Furthermore, for insider threat evaluation and mitigation, we have formulated algorithms to classify insiders into possible malicious and non malicious insiders. We have introduced a new graph called Knowledge Bayesian Attack Graph, which uses Bayesian network concepts and the knowledge graphs and the dependency graph to predict the risks associated with a user’s request to access an object. We also have devised a method for insider attack detection by profiling traceability links. This model detects an insider’s malicious activities targeted at tampering the contents of files. It uses the concept of traceability links from the software engineering field. Our approach mainly employs document dependency traceability links for insider attack detection.

**Personnel Supported:**

**Faculty:**
1. Brajendra Panda
2. Henry Hexmoor
3. Yi Hu (Northern Kentucky University – as a subcontract from the University of Arkansas)

**Research Associate:**
1. Narendra Kamila
2. Prashanth Alluvada

**Graduate Students:**
1. Qutaibah Althebyan (Completed Ph.D.)
2. Yanjun Zuo (Completed Ph.D.)
3. Yermek Nugmanov (Completed M.S.)
4. Naveen Ramanathan (Completed M.S.)
5. Seth Wilson (Completed M.S.)
6. Sandeep Bhattaram (Completed M.S.)
7. Lavanya Alapati (Completed M.S.)
8. Siddharta Gadang (Completed M.S.)
9. Frank Ching (Completed M.S.)
10. Joseph Hoag
11. Raghavan Vangipuram
12. Prafulla Kota
Resulting Ph.D. Dissertations


Resulting M.S. Theses


The above dissertations and theses can be found at the University of Arkansas library.

**Publications resulting from the Project:**


16. Rami Samara and Brajendra Panda, “Investigating the Effect of an Attack on a Distributed Database”, In Proceedings of the 7th Annual IEEE Information Assurance Workshop, West Point, NY, June 21-23, 2006. This paper was selected as one of seven "Best-Paper Nominees".


Accepted for Publication (In Press)

1. Hadi Sabaa and Brajendra Panda, “Data Authentication and the Corresponding Provenance Information Management”, Accepted for Publication in the Special Issue of the Journal of Digital Information Management (JDIM). Paper acceptance rate for this special issue was 18.6% (8 papers accepted out of 43 papers submitted).