Detecting Threatening Behavior Using Bayesian Networks

Kathryn Laskey, Ghazi Alghamdi, Xun Wang
George Mason University
SEOR Department
Fairfax, Virginia 22030
703-993-1640
{klaskey,galghamd,xwangg}@gmu.edu

Daniel Barbará, Tom Shackelford
George Mason University
ISE Department
Fairfax, Virginia 22030
703-993-1640
{dbarbara, tshackel}@gmu.edu

Ed Wright, Julie Fitzgerald
Information Extraction And Transport, Inc.
1911 North Fort Meyer
Suite 600
Arlington, Virginia 22209
703-841-3500
{ewright, jfitzgerald}@iet.com

Keywords: Information Security, Behavior Model, Multi-Entity Bayesian Networks, Document Relevance Classification, Insider Threat Detection, Access Control

ABSTRACT: *This paper presents an innovative use of human behavior models for detecting insider threats to information systems. While most work in information security concerns detecting and responding to intruders, violations of system security policy by authorized computer users present a major threat to information security. A promising approach to detection and response is to model behavior of normal users and threats, and apply sophisticated inference methods to detect patterns of behavior that deviate from normal behavior in ways suggesting a possible security threat. This paper presents an approach, based on multi-entity Bayesian networks, to modeling user queries and detecting situations in which users in sensitive positions may be accessing documents outside their assigned areas of responsibility. Such unusual access patterns might be characteristic of users attempting illegal activities such as disclosure of classified information. We present a scalable proof of concept behavior model, provide an experimental demonstration of its ability to detect unusual access patterns in simulated situations, and describe future plans to increase the realism and fidelity of the model.*

1. Introduction

Defending against an insider who attempts to misuse his access privileges is one of the most significant problems facing network security. An authorized insider can violate a system security policy for several reasons and in a multitude of ways but not all violations are true threats. An insider's privileges may range from those of a novice user to a system administrator. An insider user can be a threat against all three computer security objectives: confidentiality, integrity, and availability. One of the key findings of the eighth annual CSI/FBI 2003 report, "Computer Crime and Security Survey" [10], is that insider abuse of network access was the most cited form of attack or abuse although 92% of the respondents' organizations employ some manner of access control mechanisms. Hence, more attention must be paid to insider users allowed access to system resources in order to reduce risks imposed by them.

In this paper we focus on a particularly insidious threat: that posed by individuals who misuse their privileges to gain access to sensitive information in order to make it available to unauthorized parties (e.g.: other states, terrorists), or to manipulate it with the purpose of producing misguiding analysis. Examples of such cases are those of Robert Hanssen [4], convicted of trading secrets to the Russians in exchange for money and diamonds, and Aldrich Ames [3] who sold secrets to the KGB in exchange for money.

Modeling user behavior can provide us with insights for understanding types of threats that normally go undetected. A user's day-to-day actions modeled over time can be used to alert a security manager to possible masqueraders, clandestine users, or misfeasors. It would be possible for a smart user to slowly change his profile over time. To respond to this possibility, historical user profiles can be compared and analysis derived to detect when a user has deviated over a period of time from a generic normal work pattern. While these observations cannot be used to conclusively show that a user has done anything wrong, they can be used to alert a security manager to a possible problem. Human behavior exhibits both systematic regularities and inherent unpredictability. Social and behavioral science research has led to improved understanding of the relationships between an individual's innate personality, values, cultural traditions, life experiences, and behavior patterns. Nevertheless, no matter how good our models become, uncertainty will remain a fundamental aspect of any problem or situation involving human behavior. For this reason, modeling tools are required that can represent and exploit systematic relationships while also accounting for uncertainty and unpredictability in human behavior. This paper demonstrates the use of multi-entity Bayesian networks to model both the systematic features and the uncertainties in user behavior over time, and to accumulate evidence to distinguish normal from threatening user behavior.

2. Background

Several articles have been published in Bayesian network applications in network security. Burroughs, Wilson and Cybenko [1] provide an analysis of distributed intrusion detection systems using Bayesian methods. The main goal of their work is to defend computer networks against attackers. Information provided by intrusions detection systems (IDSs) is gathered and divided into its component parts such that the activity of individual attackers is made clear. This approach involves the application of Bayesian methods to data being gathered from distributed IDSs in order to improve the ability to detect distributed attacks against infrastructure and preliminary phases of distributed denial of service attacks as early as possible. Bayesian multiple hypotheses tracking (BMHT) algorithms generate and store all possible hypotheses that could explain the data being measured. Every hypothesis is evaluated against the understanding of the sensor behavior and the dynamics of the target. All hypotheses must be evaluated to determine their likelihood. The hypothesis that has the greatest likelihood is assumed to be the correct one. As new information arrives, the likelihood of each hypothesis is adjusted and belief in that hypothesis is either strengthened or weakened.

In [6], a Bayesian statistical model was developed to model user behavior where invalid user behavior is determined by comparing user current behavior with their typical behavior and comparing their current behavior with a set of general rules governing user behavior formed by system administrators. This prediction model has provided results that are very close to the actual user behavior with obvious similarities between results and actual data. The results were improved after applying intervention mechanisms.

Hierarchies of dynamic Bayesian network models, described in [5], were developed to compute the likelihood of various cyber attacks by dynamically adding evidence to the networks and solving the implied probability equations with a Bayesian network solution algorithm. Security situation assessment and response evaluation (SSARE) [11] provides understanding and timely management of rapidly changing cyber battle space through the application of dynamic, knowledge-intensive, Bayesian and decision-theoretic methods. It dynamically composes models in a data-driven way to develop situation-specific hypothesis to respond to the central task of cyber command and control.

3. Our Methodology

3.1 Bayesian Networks

Bayesian probability theory is a powerful technology for constructing models of phenomena involving uncertainty. Probabilities express degrees of plausibility or likelihood on a scale ranging from certainty through impossibility. Bayesian models can combine expert knowledge with observational data, and can be refined over time through learning from observation. Recently, a powerful new set of modeling methods has emerged that combine graph theory with Bayesian probability, enabling the construction of highly complex models involving large numbers of interrelated hypotheses. A *Bayesian network* encodes a probabilistic model over a set of related variables by using a directed graph to represent qualitative relationships and local probability distributions to represent quantitative information about the strength of the relationships. Bayesian networks can represent both causal and statistical dependency relationships. Figure 3.1 shows a simple Bayesian network representing user

behavior for a document retrieval task.[1] The figure shows a set of *random variables* representing uncertain hypotheses. The random variable *GlobalIntention* represents whether a user is normal or a security threat. The probability that a user is a threat is influenced by the value of the *Motive* random variable. Although users are likely to be normal regardless of motive, users with personal, financial or political motives are more likely to be threats. The remaining random variables represent the user's assigned task, the task for which the user is performing a given query, and relevance ratings of a retrieved document with respect to each of the tasks. Each random variable has a set of mutually exclusive and collectively exhaustive possible values, and a set of *local probability distributions* that specify the probabilities of its values given the values of its parents.

A Bayesian network such as that shown in Figure 3.1 and 3.2 can be used as a generative model to simulate user behavior or as a recognition model to infer unknown user characteristics and future user behavior from known user characteristics and past user behavior. Figures 3.1 and 3.2 illustrate use as a recognition model for two different patterns of evidence. In Figure 3.1, the document is rated highly relevant to the assigned task and highly non-relevant to the other possible tasks, thus reinforcing the prior expectation that a normal user is performing a task-relevant query. In Figure 3.2, the document is rated highly irrelevant to the assigned task and highly relevant to a different task, substantially increasing the likelihood of a task-irrelevant query and also increasing the likelihood that the user is a threat. Note, however, that the probability of Threat remains low. It takes more than a single questionable document retrieval to arouse serious suspicion. Threats are identified from patterns of user behavior that occur over time. To model such patterns requires a more expressive modeling technology than standard Bayesian networks.

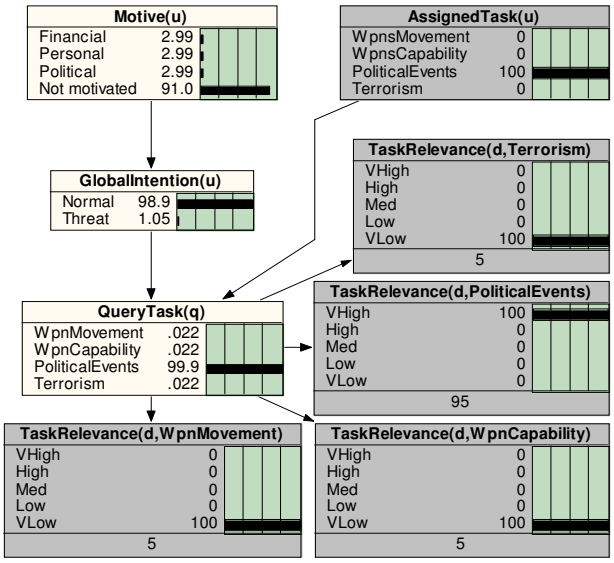

Figure 3.1 Task Relevant Document Model

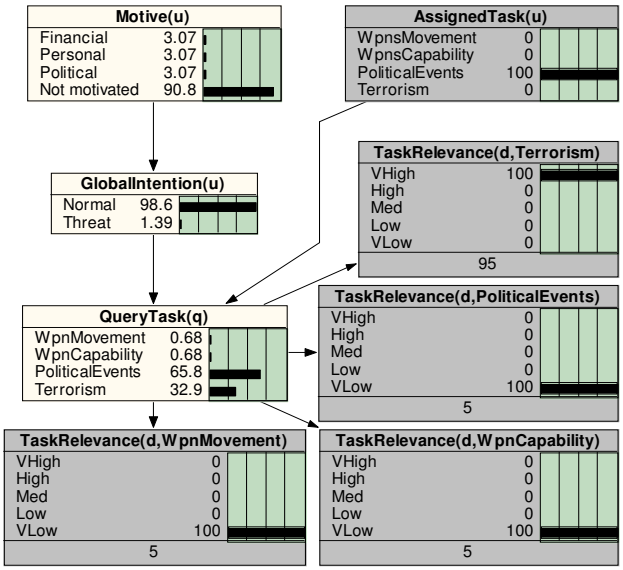

Figure 3.2 Non Task Relevant Model

3.2 Multi-Entity Bayesian Networks

Standard Bayesian networks are limited to problems in which the same set of random variables applies to all problem instances, and only the evidence is different from problem to problem. A much more flexible representation capability is required to model human behavior in all but the most stereotypical situations. *Multi-entity Bayesian networks* (MEBNs) expand upon

[1] Screen shots of Bayesian networks are from the Netica® Bayesian network package.

standard Bayesian networks in their ability to encode repeated, parameterized argument structures called *MEBN Fragments* (MFrags). An MFrag is a modular component representing a fairly small, separable, and conceptually meaningful part of the total argument structure supporting or denying a given hypothesis. MFrags can represent alternative hypothetical world states, evidence that bears upon which hypotheses are true, and chains of argument relating evidence to hypotheses. MFrags can be combined to build models relating complex configurations of many features, and can be re-used in multiple scenarios. Figure 3.3 shows how the model of Figure 3.1 and 3.2 can be represented as a set of MFrags. Each MFrag has a set of *resident random variables* (shown in white) whose local distributions are defined in the MFrag, *input random variables* (shown in light gray) whose values condition the local distributions of the resident random variables, and *context random variables* (shown in darker gray), which must have value *True* for the local distribution defined in the MFrag to apply. The random variables take arguments called *entities*. For example, the query task MFrag of Figure 3.2 applies when the entity *u* (representing a user) is equal to the value of *PerformingUser*(*q*) for the entity *q* (representing a query) – that is, when user *u* is performing query *q*. It specifies the relationship between the user's assigned task, the user's intention, and the query task.

The local distributions of MFrags can be learned from observations as evidence is accrued. MFrags can represent complex models involving multiple actors, multiple documents, and multiple computer systems. They can appropriately handle the resulting complex correlations both in inference and in learning. The learned MFrags are applied at runtime to detect anomalous events in a process called situation-specific Bayesian network construction. As evidence arises, suspicious configurations trigger "suggestors" which bring in MFrags that provide potential explanations for the normal as well as anomalous patterns. The constructed situation-specific Bayesian networks are then used to infer the probability of hypotheses of interest, and to trigger alerts as necessary.

MEBN theory has been implemented in Quiddity*Modeler (Q*M), part of a suite of Bayesian Inferencing tools developed by Information Extraction and Transport, Inc. (IET). Q*M is a knowledge representation language based on frames (a widely used knowledge representation in Artificial Intelligence) augmented in various ways to express uncertainties. In addition to frame (class) abstractions organized by "is-a" hierarchies inherited from the frame system, Q*M supports mechanisms to express uncertainties about the value of variables, the reference to instances, the existence of instances, and the type of instances. Q*M allows for expressing domain knowledge as pieces of BNs, called Bayesian Network Fragments or MFrags, in a modular and compact way, facilitating reuse. Instances of probabilistic frames are created dynamically for each instance, allowing situation-specific probabilistic inference.

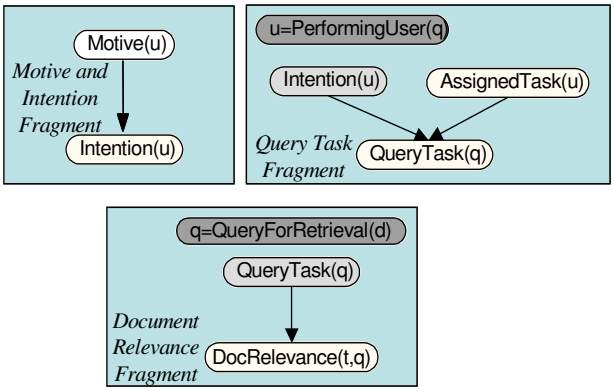

Figure 3.3 MEBN Fragments for Document Retrieval

4. Our Model

Our current model consists of seven MFrags that model queries and document accesses performed by users. Within that constraining scenario, we identified information useful for distinguishing normal users from those who pose a threat. The model is implemented in IET's Q*M frame language. A frame is a software object representing a type of entity. Each frame contains slots that represent attributes of entities of the associated type. Slots contain information used to determine the probability for each slot. At run-time, instances of each frame are created to represent particular entities of the type modeled by the frame as necessary.

Figure 4.1 shows how instances of the MFrags of our model can be assembled into a unified Bayesian network. This network is composed of nodes in parent-child relationships. Each node has several states, representing the slot-values of slots appearing in frames. We will discuss each MFrag, explaining its structure and role.

User—The user MFrag represents an individual user's profile. The slots included are name, clearance level, role, motive, intention, assignment, and other activity. Motive, intention, assignment and other activity slots are used to determine threat and these are explicitly represented in the network shown in Figure 4.1. Most user slots are reference slots, pointing to other MFrags,

but it is useful to have a single user reference fragment for use throughout inferencing.

User Background—We currently model three areas that may serve as indicators that a user is likely to be a threat: political activities, personal background, and financial background. In each area we make fairly coarse distinctions such as serious concerns, minor concerns, and no concerns. In future work, we may flesh out the relevant user background, improving how we determine the threat level posed by each user and making finer distinctions regarding the levels of concern. While our model includes variables that represent user background we are not using this background information as evidence in our model.

User Assignment—Each user has an assigned geographic region and an assigned task. This information is taken as known and set as evidence on the network. We also assume that assignments are constant within a given time step. We have created fictional regions and tasks for our proof of concept network but we could easily substitute real regions and tasks. Given that the assignment nodes have no parents, we can easily increase the numbers of both regions and tasks as necessary.

User Intention—We currently classify users' intentions as either "normal" or "threat." We assume that users' intentions may change over several sessions. We assume users have global intentions that do not change over the time interval represented by the model and session intentions that may change from one login session to the next. The session intention is influenced by both the global intention and the previous session's intention. We do not directly know either the global or session intention of users. The global intention is currently influenced by user motive. The session intention influences the user query nodes and thus can be inferred from the pattern of documents accessed over time.

User Other Intention—Our goal is to build a model that not only identifies malicious users but also indicates the nature of the potential threat. Currently we model the threat by attempting to identify an information source the malicious user may be trying to identify, as well as any regions and/or tasks in which the user is showing interest in addition to his or her assigned region and task. The user's other intention also influences the documents he or she accesses, and can be inferred from the pattern of document retrieval behavior.

Document—Documents have sources (providers of information contained in the documents) and region and task classifications. Documents are rated to provide a measure of how relevant each document matches each of task and region. The relevance rating is currently provided manually but in the future relevance will be assigned according to a document relevance classification system. Techniques to measure document relevancy such as the ones described in [2] will be used for this purpose.

Query—Users perform queries that result in document accesses. We assume for any given query that the user is seeking information about a source, a task and a region. The query source represents the source that provides the intelligence information in a retrieved document, as opposed to the author of a document. If a user is attempting to identify a given source, then whether or not the user explicitly queries on the source, task or region, the user is likely to access documents citing this source. A query is also likely to return a document relevant to the query task or region.

Figure 4.1 shows the seven MFrags assembled into a unified model for a single session, a single query, and a single document access. The *context* variables in the upper left corner represent relationships that must hold among the referenced entities in order for the relationships in the constructed model to hold. The context variables state that the model refers to a user *u* who accessed a document *d* retrieved in response to a query *q* performed by *u* during a session *s*. The other random variables in the model represent the values of slots in the frames.

For the network shown in Figure 4.1, the user's assigned task and region have been set as evidence, as well as the results of a query (as shown by the evidence set on both task and query relevance). Even though the user is, in this example, querying for documents outside of his/her assigned task and assigned region, our belief that the user is a threat increases only by a small amount. The prior probability for a user being a threat is 1.08%. After setting the evidence used for this example, the probability has increased to 2.83%, as reflected in the *GlobalIntention(u)* node. This minor shifting of our belief is in keeping with the fact that the great majority of users are not threats; that users who are threats will act like normal users the majority of the time; and that detection of threats requires the accrual of data over time as we do using dynamic nodes.

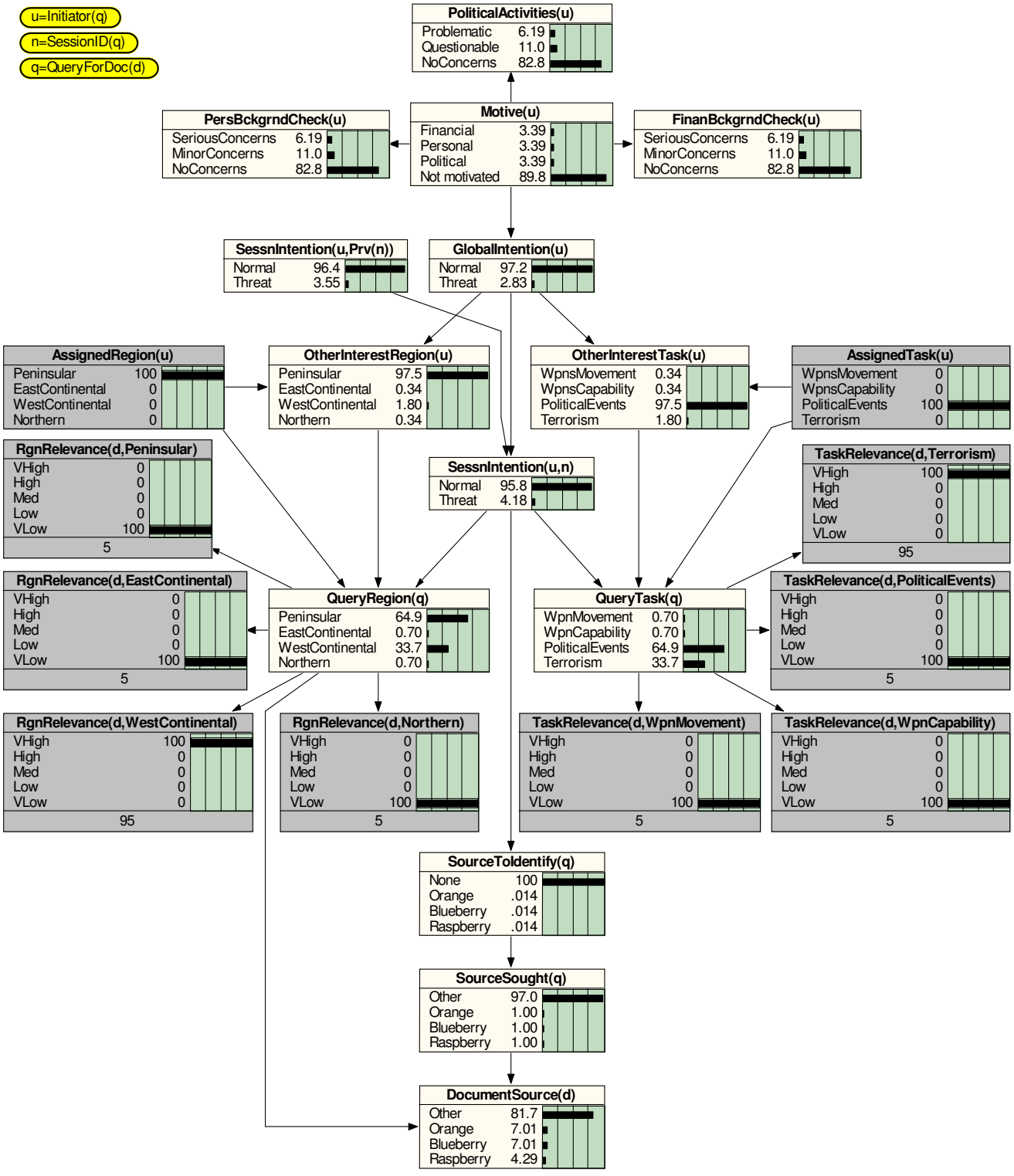

Figure 4.1 Proof of concept model for detecting insider threats, with evidence set

MEBN logic as implemented in the Quiddity*Modeler language allows many instances of each frame to be combined automatically into a single complex model involving multiple users, multiple sessions per user, multiple queries per session, and multiple document retrievals per query. Quiddity*Modeler is also powerful enough to allow us to model access sequences and patterns. This is a planned extension to the current models and is discussed below. In the experiments described below, the Bayesian network for a single session contained hundreds of nodes, and the experiment was carried out over 100 sessions.

5. Simulation Experiment

5.1 Experiment

The aim of our experiment is to distinguish users' type (normal/threat) through a set of actions by the users' (multiple document queries) over a period of time (multiple sessions). We implemented our experiment by creating two identical Bayesian networks operating simultaneously. The *ground truth* network was used as a generative model to simulate a user's intention and behavior. The model simulated session intention, a set of queries for the session, and a set of document retrievals and associated relevance ratings. The *inference* network was used as a recognition model to detect threatening user behavior.

In a specific session, the two parallel models are the same (same nodes, same arcs between the nodes and same belief tables for the nodes). Each experimental session had a global intention (normal or threat), an assigned region, and an assigned task set by design. The values of the design random variables e *GlobalIntention*(u), *AssignedRegion*(u), and the *AssignedTask*(u) were applied as evidence in our generative model. We then sampled the random variables *OtherInterestRegion(u)* and *OtherInterestTask(u)* and applied the sampled values as evidence in the ground truth network. We also set the values of *AssignedRegion*(*u*) and *AssignedTask*(*u*) as evidence in the inference network, because it is assumed that their values will be available to the inference system.

Finally, for each session we simulated a set of queries and document accesses, applied the simulated document relevance ratings and sources as evidence in the inference network, and computed the posterior probabilities for *GlobalIntention*(*u*):

P(GlobalIntention=Normal | data, assigned task/query)

P(GlobalIntention=Threat | data, assigned task/query)

This process was repeated for each of 100 sessions for each user. For a specific user we obtained a sequence of probabilities, one for each session, allowing us to compare the inferred results with the ground truth.

5.2 Simulation and Inference

We simulated the behavior of a total of 192 users, distributed in the following manner: 16 Normal users with 5 queries of 4 documents per session for 100 sessions; 48 Threat users under the same conditions; 16 Normal users with 2 queries and 10 documents per session for 100 sessions; 48 Threat users under the same conditions; 16 Normal users with 10 queries and 2 documents per session for 100 sessions; 48 Threat users under the same conditions. In each case the procedure was as follows:

1. Use Quiddity*Modeler frame language to create the generative and recognition BN models for a given design case (user type, assigned region, assigned task, number of queries per session, number of documents per query).

2. For each ground truth user type (16 Normal users and 48 Threat users) do the following procedures:
   2.1. Set the values of *GlobalUserIntention*(*u*) to the design value (Normal/Threat) in the generative model. Set the values of *AssignedRegion*(*u*) and *AssignedTask*(*u*) values to the design value in both the generative and inference models. Then we sample the random variables *OtherInterestRegion*(*u*) and *OtherInterestTask*(*u*) from our generative BN model and set the sampled values as evidence in our generative model.
      2.1.1. For each of 100 sessions do
      2.1.2. Simulate data from the generative model for each query in the session and each document access for the query, where the number of queries and the number of document accesses is set by design
      2.1.3. Set the documents relevance ratings and source information as generated by the ground truth model as evidence in the inference model.
      2.1.4. Use our inference BN model to compute the two posterior probabilities: P(GlobalIntention=Normal | data, assigned region/task);

P(GlobalIntention=Threat | data, assigned region/task).

2.1.5. Advance both our generative BN model and inference BN model to the next session.

6. Results

Figures 6.1 through 6.4 show typical results of experimental runs for a normal user, two detected threats, and an undetected threat. For each of 48 test simulations of normal users, the probability of *Normal* remained near 1.0 over all 100 sessions. Of 144 test simulations with threat users, 14 (about 10%) were indistinguishable from the plots of normal users. An additional 2 *Threat* users had high probabilities above 80% *Normal* over all 100 sessions, but noticeably lower than for normal users (e.g., around 90%) for some set of sessions. For the remaining 128 *Threat* users (about 89%), the probability of *Normal* dropped to single digits. In most cases, the probability of *Normal* remained low, although in a few cases it began to climb at some point. That is, according to this model, document retrieval patterns alone would be sufficient to definitively detect almost 90% of the threats in 100 sessions. Approximately 1% of simulated *Threat* users showed patterns only slightly, although noticeably different from normal users. Below we include four typical plots that show the inferred probability of "Normal" and "Threat" over 100 sessions.

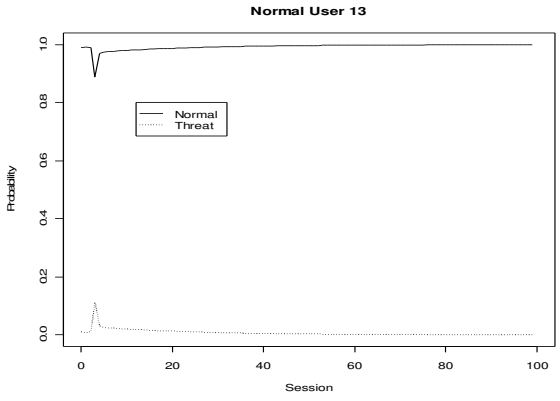

Figure 6.1 Typical plot for a normal user

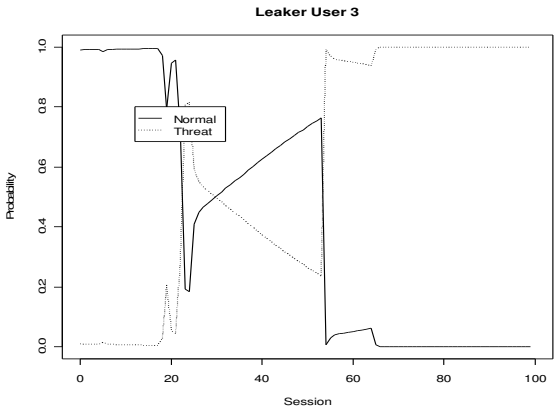

Figure 6.2 Typical plot for a threatening user that is detected after some rounds of suspicious behavior

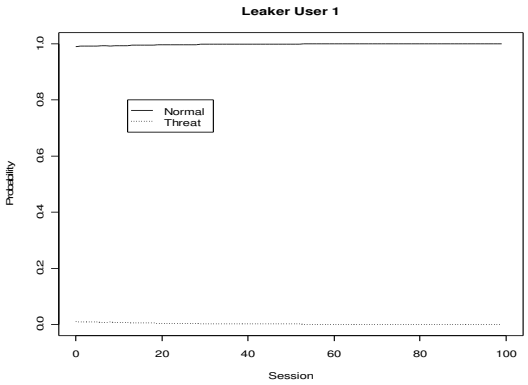

Figure 6.3 Typical plot for a threatening user who is undetected

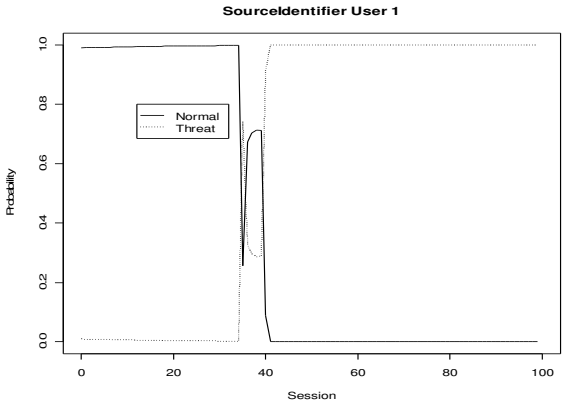

Figure 6.4 Typical plot for a detected threatening user

7. Discussion and Future Work

The model presented here illustrates the concept of detecting insider threats from document retrieval patterns based on a model of user behavior. The model has not been validated against actual user behavior. The probabilities used in the model were assessed judgmentally and are not reflective of actual measured user behavior. Our experiments have demonstrated the ability of a behavior model such as this to perform reasonable inferences using data generated from the model itself. A definitive demonstration of the value of this modeling technology for information security applications would require learning the parameters of the model from field data and testing the model against field data not used to train the model. Nevertheless, this simple model demonstrates the potential utility of MEBN technology for modeling user behavior in information systems. Our results demonstrate that user behavior modeling using MEBNs is a promising approach to the problem of detecting insider threats in information systems.

One future direction of our work is to examine the problem of generating ground truth data that addresses a more realistic insider behavior. We also plan to discriminate between different types of threats, e.g., users with intentions of leaking information versus those with the intention of influencing policy making. A further improvement to the model includes integrating with other insider user attributes such as insider skill, knowledge, clearance level, access sequences and patterns, log-in time for sessions, and realistic security policies.

Another avenue of research we plan to explore is the use of data mining techniques to generate evidence for our Bayesian Network. In particular, we are interested in automatically computing the task relevance of a document that the analyst is accessing. For this task, we can borrow the ideas from [2] on computing the relevance of a document when performing focused crawling in the web. Concretely, given a pre-defined ontology of topics, where a series of topics have been marked as relevant to the task in hand, and a given document being accessed by an analyst, we are interested in computing a measure of relevance *R(d)*, as $R(d) = \sum_{relevant(c)} P(c \mid d)$. For every topic c which has been marked, *relevant(c)* is set to true. $P(c|d)$ represents the probability that *d* addresses the topic *c*. This can be computed using the ontology and a chain rule:

$$P(c \mid d) = P(parent(c) \mid d) P(c \mid d, parent(c)).$$

Using Bayes rule it is possible to express the last conditional probability as:

$$P(c \mid d, parent(c)) = \frac{P(c \mid parent(c)) P(d \mid c)}{\sum_{c', parent(c')=parent(c)} P(d \mid c')},$$

where the sum ranges over siblings of *c* in the taxonomy. Finally $P(d|c)$ can be computed using a a Bernoulli binomial distribution model:

$$\mathrm{P}(c,d) = \prod_{t \in d} \left(\frac{n(d)!}{n(d,t)! * (n(d) - n(d,t))!} \right) \theta(c,t)^{n(d,t)}$$

In this last expression, *t* represents terms (words) in the document, *n(d)* is the number of words in *d, n(d,t)* the number of times the *t* appears in *d*, and $\theta(c,t)$ the probability distribution of *t* in topic *c*. When computed, the relevance measure *R(d)* can be fed into the Task Relevance nodes of the Bayesian Network as evidence. In this way, a link between the pattern of document accesses from an analyst and our Bayesian model can be established. We plan to design experiments to test this technique in the near future.

8. Acknowledgements

Work for this paper was performed under funding provided by the Advanced Research and Development Activity (ARDA), under contract NBCHC030059, issued by the Department of the Interior. Additional support was provided by the US Navy. The views, opinions, and findings contained in this paper are those of the author(s) and should not be construed as an official position, policy, or decision, of ARDA, the Department of the Interior, or the US Navy unless so designated by other official documentation

9. References

[1] Daniel Burroughs, Linda Wilson, and George Cybenko. "Analysis of Distributed Systems Using Bayesian Methods." Performance, Computing, and Communications Conference, 2002. 21st IEEE International , 2002 Page(s): 329 –334

[2] Soumen Chakrabarty, Martin van den Berg, and Byron Dom ``Focused crawling: a new approach to topic-specific Web resource discovery,'' Computer Networks, 31(11-16), pp 1623-1640, 1999.

[3] CNN.com ``Rationalizing Treason: An interview with Aldrich Ames,'' http://www.cnn.com/SPECIALS/cold.war/experience/spies/interviews/ames/

[4] CNN.com ``The Case Against Robert Hanssen.'' http://www.cnn.com/SPECIALS/2001/hanssen/

[5] Finn V. Jensen, "Bayesian Networks and Decision Graphs", Springer-Verlag 2001.

[6] J. Pikoulas, W. Buchanan, M. Mannion, and K. Triantafyllopoulos. "An Intelligent Agent Security Intrusion System." Engineering of Computer-Based Systems, 2002. Proceedings. Ninth Annual IEEE International Conference and Workshop, 2002. Page(s): 94 -99.

[7] Wright, E.; Mahoney, S.; Laskey, K.; Takikawa, M.; Levitt, T.; "Multi-entity Bayesian networks for situation assessment," Proceedings of the Fifth International Conference on Information Fusion. Volume: 2, 2002. Page(s): 804 -811 vol.2

[8] Judea Pearl, "Decision Making Under Uncertainty". ACM Computing Surveys, Vol. 28, No. 1, March 1996.

[9] Result of a Three-Day Workshop: *Research and Development Initiatives Focused on Preventing, Detecting, and Responding to Insider Misuse of Critical Defense Information Systems*. August 1999.

[10] The Eighth Annual CSI/FBI 2003 report: "Computer Crime and Security Survey."

[11] D'Ambrosio, B., M. Takikawa, J. Fitzgerald, D. Upper, and S. Mahoney. "Security situation assessment and response evaluation (SSARE)", Proceedings of the DARPA Information Survivability Conference & Exposition II, Volume I, IEEE Computer Society, June 2001, pp. 387-394.

10. Author Biographies

GHAZI ALGHAMDI is an officer in the Royal Saudi Navy Forces. He is pursuing his Ph.D. degree in Information Technology at George Mason University as part of the training scholarship programs provided by the Royal Saudi Naval Forces through the training agreement with the United States Navy. His research concentration is the application of Bayesian Networks in information security.

DANIEL BARBARA is a Professor of the Information and Software Engineering at George Mason University. His research interests are in Data Mining and its applications to a variety of domains.

JULIE FITZGERALD is a scientist with Information Extraction and Transport Inc. Her research areas include evaluation of large artificial intelligence systems, knowledge representation, and Bayesian inferencing.

KATHYRN LASKEY is an associate professor of Systems Engineering and Operations Research at George Mason University. Her research emphasis is Bayesian approaches to knowledge representation and inference. She has applied multi-entity Bayesian networks to a variety of problem domains.

TOM SHACKELFORD is a Computer Scientist for the Joint Interoperability Test Command, Fort Huachuca, AZ. He is pursuing his Ph.D. in Information Technology at George Mason University as part of the Information Assurance Scholarship Program.

XUN WANG is a PhD student of System Engineering and Operation Research at George Mason University. He is interested in Bayesian statistical models and inferences.

ED WRIGHT is a scientist with Information Extraction and Transport, Inc. His research areas include intelligence data fusion, Bayesian inferencinging, and modeling complex problem domains.