Towards a Cyber Ontology for Insider Threats in the Financial Sector

Gokhan Kul* and Shambhu Upadhyaya
State University of New York at Buffalo, Buffalo, NY, USA
gokhanku@buffalo.edu, shambhu@buffalo.edu

Abstract

Insider attack has become a major threat in financial sector. Currently, there is no insider threat ontology in this domain and such an ontology is critical to developing countermeasures against insider attacks which are very serious and pervasive security problems. In this paper, we offer a methodology to categorize insider attack suspicions using an ontology we create, which focuses on insider attacks in the banking domain targeting database systems. The scheme we propose takes a suspicion alert as input that triggers the ontology mechanism to analyze the chronology of the events. Our model formulates the ordinary processes that take place in a financial organization and systematically evaluate events in a sequential order. To create the ontology, we use a top-down analysis approach to define a taxonomy and identify the relationships between the taxonomy classes. The ontology is mapped onto the Suggested Upper Merged Ontology (SUMO), Friend of a Friend (FOAF) and Finance ontologies to make it integrable to the systems that use these ontologies and to create a broad knowledge base. It captures masquerade, privilege elevation, privilege abuse and collusion attacks and can be extended to any other novel attack type that may emerge. It classifies an attack using the knowledge base provided and the missing relationships between classes. We validate the ontology showing how description logic works with a given synthetic scenario which is created by banking experts.

Keywords: Cyber ontology, financial sector, relational database systems, taxonomy

1 Introduction

Most security systems are built to protect sensitive information from external threats. Networked systems and information technology systems are changing with rapid innovations and they have been claiming more crucial roles in critical infrastructures. This rapid development has made the issue of information security more apparent due to the information contained in these systems.

Nowadays, insider threat is becoming an extremely serious security problem due to the threat it poses to the monetary assets and the sensitive customer data for financial institutions. The RAND report [1] addresses insider threat as "malevolent (or possibly inadvertent) actions by an already trusted person with access to sensitive information and information systems." If there are not enough precautions taken, it can lead to insider attacks.

The 2014 U.S. State of Cybercrime Survey [2] emphasizes the severity of insider attack. According to the survey, 37% of organizations have experienced an insider incident, and 32% of the respondents to the research conducted say that the attack perpetrated by insider is more damaging than outsider attacks. In 82% of these cases, private or sensitive information was unintentionally exposed, in 76% of incidents, confidential records were compromised or stolen. In 71% and 63% of these incidents, respectively, customer and employee records were compromised or stolen. This data only includes the

Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, volume: 6, number: 4, pp. 64-85

*Corresponding author: Department of Computer Science and Engineering, State University of New York at Buffalo, 338 Davis Hall, Buffalo, NY, 14260, Tel: +1-(716)645-3180, Web: http://www.cse.buffalo.edu/~gokhanku/

known insider incidents. However, it is expressed in this report that 28% of electronic crime events are known or suspected to have been caused by insiders and in 46% of electronic crimes, insider attacks were more costly or damaging to their organization. The report also shows that 75% of cases were handled internally without legal action or law enforcement, mostly because of lack of evidence or not enough information to prosecute. Only 10% of cases were handled internally with legal action and 12% of the cases were handled externally with notification of law enforcement while only 3% of cases were handled externally by filing a civil action. The percentage of cases that could not be prosecuted due to lack of evidence raises questions on the reliability of security systems toward identifying insider threats.

This paper is the extended version of the work presented in [3]. The banking domain differs from the others in the sense of information security. Most of the employees like tellers and customer representatives in a banking institution have the ability to access very sensitive information. This makes performing an insider attack in banking organizations easier. An attack that is coming from an employee within an organization can go unnoticed for a very long time [4] and the implications can be very costly.

Examining the chronology of events is not enough to understand if the intent is malicious. Insiders usually have authorized access to the information they target, and their behaviors are very difficult to distinguish from normal activities, which leads to false negatives. On the other hand, a benign intent can sometimes be observed as malicious behavior. Acting upon a false positive without evaluating the sequence of events and their details carefully can lead to a message of distrust between the employees and the administration in an organization.

We focus on insider attacks on relational database management systems for a variety of reasons. First, keeping the focus on a specific but important domain allows us to contain the scope of the model at a more manageable level. Second, even though there are other data preservation techniques and systems, relational databases are heavily used in back-end servers to store financial data, which consists of a lot of sensitive information. This makes relational databases a primary target for cyber crime and identity theft. The aim of this effort is to develop an ontology of this area, expressed in the Web Ontology Language (OWL) that ensures integration with other knowledge domains and enables data integration across different data sources. Semantic web applications are becoming more popular by the day and ontology is the most important enabling technology of these applications. Basically, it describes terms and different relationship types between terms. In this paper, we create a taxonomy of insider threats and identify the relationships between the entities we define in the taxonomy. These entities and relationships are used to create an insider threat ontology which is then mapped onto upper ontologies and domain ontologies that are commonly used in financial systems. We use this technology to make background knowledge of the domain explicit for computers, so that for context we feed, the system can "understand" and "reason with" the information as humans do.

The contributions of this paper are providing support to the security experts in the investigation process; creating a framework of a cyber ontology for insider threats in the financial sector focusing on relational database management systems; integrating this ontology with commonly used ontologies the Suggested Upper Merged Ontology (SUMO) [5, 6], Friend of a Friend (FOAF) [7, 8] and Finance [9] to make it applicable and integrable to the systems that use these ontologies; and finally, exploiting the capabilities of ontologies, providing a methodology to distinguish between benign and malicious event when a suspicion arises.

There are several challenges when developing such a framework. Mainly due to regulatory reasons, financial institutions cannot share sensitive information which includes their internal structure. This results in uncertainty on analyzing what data is available to us and shaping how to validate our initial proof of concept. The other major challenge is correctly building the ontology. The representation of the world should be accurate and precise. Inaccurate and insufficient analysis of the domain creates risk of misrepresenting the domain.

Section 2 reviews related work in the literature. Section 3 starts with describing the methodology of

the overall system to identify insider attacks, and gives the details of creating the taxonomy and ontology on insider threats. Section 4 explains the required information to create the knowledge base and for the evaluation of the ontology. Section 5 performs the validation of the ontology based on insider attack scenarios given. Section 6 discusses the advantages and contributions of our research, and finally Section 7 presents the future work.

2 Related Work

The prior research in this area is along two different directions. One considers psychological aspects and the other considers physical aspects of insider threats. The research approach taken, however, can be categorized as technical, social or socio-technical [10]. The phrases "insider" and "insider threats" are terms that have ambiguous definitions, but are known to many for what they mean. Most of the research which has been done on insider threats is mainly on the psychological structure and incentives of these attacks and how to prevent them on general cases. There are cases in which researchers have focused on physical threats from insiders. For example, the work performed on [11] focuses on how to protect data centers from physical attacks and insider threats. Some instances may include a recently terminated employee, a user on a computer that is logged in, or even a janitor. No matter who the insider is, the potential threat to an organization is a problem that many organizations need to account for.

Hunker and Probst [12] go into detailing what exactly an insider and insider threat are, while giving examples of solutions to the problem of insider threats. They give definitions for "insider", "insider threat", as well as detailed issues that arise when managing insider threats and the lack of data on the topic while describing the multiple approaches to an insider attack. They describe the technical and socio-technical approaches to dealing with insider attacks and discuss the sociological, psychological, and organizational approaches to dealing with insider attacks. The authors explore the wide range of different people that can be insiders and they describe all of the aspects that go into making someone an insider. This level of detail is carried into the description of an insider attack, showing how there can be different types including accidental threats as well as malicious attacks.

Mundie *et al.* takes the initiative to create an ontology framework for insider threat research [13]. They focus on standardization of the terms insider and insider threat while investigating the relationships between these terms. Their main aim is to provide a better understanding of the conceptual model of the insider threat, hence eliminating the inter-study differential and facilitating a standardization of terms.

Mathew *et al.* [14] state that insider attacks pose a serious threat due to the fact that current security systems are aimed at prevention of unauthorized access. They focus on the fact that not only can threats come from trusted entities within an organization, but a successful attack may be the result of multiple entities working together, termed insider collusion. Therefore, this is said to justify a call for monitoring and detection methods which take into account these potential interactions between entities. From here, they go on to detail the use of a new system, called Information-Centric Modeler and Auditor Program (ICMAP), which generates Capability Acquisition Graphs (CAGs) to represent information about physical locations of data, difficulty of access to components of the data system, etc. This graph allows for feasible analysis of which paths to insider abuse targets are the least difficult to traverse. The CAG holds information about the potential difficulty of accessing certain nodes in the system, and can therefore determine the path of least resistance. This can allow for security analysts to bolster the defenses of the systems along that path, or simply to monitor activity along these nodes for suspicious behavior. It is noted that the cost of creating, updating, and analyzing a CAG is considerably high and thus impractical to maintain in real-time. The proposed solution is to only update the CAG periodically (termed "CAG milestones"), as well as search for paths vulnerable to attack using a greedy algorithm that may not give the absolute most vulnerable path in the system, but is likely to after a number of runs. They provide

an example of a situation in which a collusion attack could be carried out undetected, with malicious activity performed under the guise of being legitimate work tasks. Therefore, such a scenario would be difficult to catch in the act. However, a CAG generated by ICMAP can trace the means through which somebody with only public access could obtain information with top-secret security restrictions.

In their paper, Costa *et al.* [15] detail the creation of an ontology for use in describing the indicators of insider threats. The primary reason cited for focusing on this area is that it had been uncommon for information about these insider threats to be circulated outside of the businesses that were typically subject to them; without a standardized method to abstract the data, doing so would have meant releasing confidential data related to the attack. Without public circulation of this information, progress in determining methods to prevent these insider threats has been severely hampered despite increasing focus in this area of research. The ontology was developed with the aid of over 800 cases of malicious insider activity compiled from various sources, all of which were natural language descriptions of the incident. These cases were analyzed using a semi-automated method which had output relationships between common concepts which were used as the basis of the classes for the ontology. The top-level classes used were "Actor, Action, Asset, Event, and Information." The ontology can then describe scenarios by showing the relationships between subclasses of these top-level classes. The paper goes on to give a series of examples for how to use the ontology to further the field of insider threat detection. While it starts by restating the usefulness of this level of abstraction for publicizing information related to threats without also disclosing organization-sensitive information, the paper also goes on to note that the semi-automation of data collection that the ontology implementation paves the way for others to develop detectors for indicators of insider threats. Also, the paper states that it would be possible for this work to be extended such that event logs (and other operational data) as well as information that organizations keep about insiders that is not as a result of direct interaction with information technology (human-resources data) could be translated and parsed in order to automatically create ontology individuals. If these processes are automated, then this would make it possible for a semantic reasoner to be constructed to classify insiders as instances of subclasses within the ontology, which would provide a clear view of specific indicators of threats. Ultimately, the development of this ontology appears to be a valuable stepping stone to further progress in the area of insider threat detection, but its greatest benefits will be lost if it is not widely used in a standard form.

The ontology schemes proposed in [13, 15] differ from the methodology introduced by us in [3]. Our methodology aims to ensure the integration with other knowledge domains to enable data integration and it models the financial domain while including a basic overview model of insider threat. The ontology modeled can be used to systematically evaluate any insider threat detection schemes in a realistic way and discover attacks that share similarities with previously identified attacks. However, this scheme assumes that there is historical data on insider attacks caught.

3 Methodology

Taxonomy and ontology are two common terminologies that are being used in information management and there are cases where people treat them as synonyms.

The term "taxonomy" could refer to a hierarchical classification or categorization system, or to an organization of concepts of knowledge, as well as a knowledge organization system designated to include term lists and classifications [16]. Except for some rare cases, defining the relationships between entities is not a concern when defining taxonomies, other than a hierarchical relationship between entities.

The term "ontology" but for its philosophical meaning, is a formal framework to represent knowledge in computer and information science. Ontologies define classes, properties of these classes and relationships between these classes within their domain. Using the relationships, we can extract other

information from these information entities and use them to identify other previously unidentified relationships between them. The authors of [17] classify taxonomies as linguistic/terminological ontologies. However, taxonomies can also be used to define ontologies when the relationships between the classes are defined and a formal structure of an ontology can be constructed with them. How to develop an ontology is summarized in [18] as

- Defining classes in the ontology
- Arranging the classes in a taxonomic hierarchy
- Defining slots and describing allowed values for these slots, namely, creating properties of the classes
- Filling in the values for slots for instances.

We model the normal world, not the insider threat, while creating the ontology framework for this study, and use normality to form the story of the attack. As mentioned in Section 1, we capture behavior anomalies, using the SQL queries that are issued on the databases of the organization, which is not in the scope of this work. By profiling user behavior in an RDBMS to detect anomalies, we can mark some queries as suspicious [19] and trigger other mechanisms to investigate what the real intent is. As shown in Figure 1. the scope of this work is within the circle. When we capture a suspicious intent, we fill the ontology slots for instances with the event logs that are also stored by the organization. These logs include most of the behaviors that can be classified as normal behavior, but exploiting the capabilities of ontologies, we can query the ontology and capture the anomalies, so that we can classify the suspicious intent. The suspicious intent can be classified as benign, or it can be classified as one of the insider attack types. The duty of the management of the organization or the security expert is inspecting the instances and connections in the ontology when it decides that it is an insider attack, and then building the story or with reverse engineering techniques.

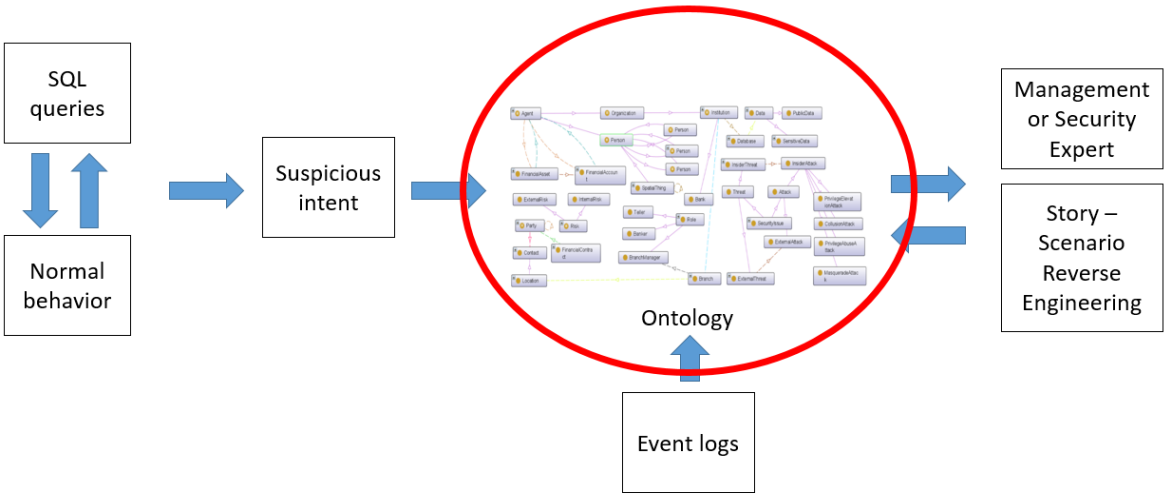

Figure 1: Insider threat identification scheme

This section identifies the methodology employed in the taxonomy and ontology development process and explains the details of the construction of ontology classes.

3.1 Ontology development

The ontology development process we employed in this paper is a top-down analysis which requires understanding the semantics of the end-users who will actually use the resulting ontology. It starts with creating a list of terms which will be used to construct the taxonomy of the structure. The taxonomy needs to include the terms that define the classes in the domain. The taxonomy needs to be limited with what the resulting ontology will cover, what will it be used for, and what types of questions the ontology will answer to. Following the creation of the taxonomy, and the hierarchy within the taxonomy, the properties of the classes should be defined along with the relationships between classes.

The validation of the ontology structure is performed through competency questions. These questions assure that the targeted value of the structure is achieved. They serve as procedures that indicate when the ontology development is sufficiently complete. The competency questions aim to ensure that the results are accurate, sufficient, and has the right level of granularity which is identified by the subject matter expert. They also ensure that the scope of the ontology is still within the limits.

It is essential to integrate the ontology created with other ontologies, as it will integrate the domain with the rest of the world. Considering that ontologies are a web of knowledge, integrating the ontology with other ontologies will create a bigger knowledge base and extend the opportunities of integrating this ontology with the existing systems. However, to increase data and information quality within a domain, we need to create an ontology that can represent that domain successfully, and creating an ontology requires expert knowledge within that specific domain as well as the skills required to create it. To create an ontology, ontology developers and domain experts need to work together. Ontologies that are created by people who lack either expert knowledge or ontology development skills may result in serious problems and wrong results. However, not all research projects have enough resources to hire people who have these skills. Also, even if the resources are sufficient, project teams may not think it is necessary.

3.2 Taxonomy

The efforts we have put into creating a taxonomy on finance domain has resulted with the taxonomy shown in Figure 2. As a result of the top-down analysis we performed with the domain experts of our collaborator banking institution, we have identified the taxonomy classes based on basic scenarios given in Section 5. The validation of these classes is made through mapping between classes and instances gathered from the mentioned scenarios. Some of the main class definitions are given in Appendix A.

3.3 Ontology

There are several types of ontologies that we can base the rules of our ontology framework.

Upper level ontology: The ontologies that belong to this level describe concepts that are the same across all knowledge domains which provides a high level of semantic interoperability.

Domain ontology: The ontologies that belong to this level describe concepts in a specific field or in a part of the world. This specific field or part of the world represents the domain that the ontology describes. Since the concepts belong to the domain, they may or may not be compatible with a concept that has the same name in a different domain ontology.

Hybrid ontology: The ontologies that belong to this level describe concepts that can be both mentioned in domain and upper level ontologies. Especially by working on integration of different systems together, the hybrid approach makes it easier to work with multiple ontologies. Some concepts can be defined universally but some concepts are described according to the domain related limitations.

Our goal is to provide a web of knowledge by integrating commonly used upper ontologies into our ontology. To achieve this task, we create a domain ontology on insider attacks focusing on financial

Figure 2: Ontology classes created from initial terminology

sector, and then we identify some ontologies that are commonly used by academia and industry that may possibly have similar classes that we identified in our ontology.

Friend of a Friend (FOAF) ontology [7] [8] describes people, their activities and relationships between each other and other objects. It allows groups of people to create social networks, which we are using to describe the relationships between customers, bank personnel and roles and hierarchy within the organizations. The common terms that we import from this ontology are "Organization" and "Person" classes as can be seen in Figure 3. After importing these classes, we expand this terms with the domain specific subclasses, to define the banking environment.

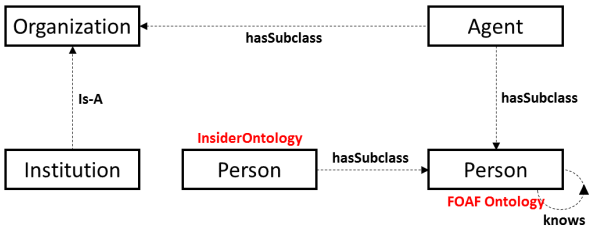

Figure 3: Integration of FOAF ontology classes

The Suggested Upper Merged Ontology (SUMO) [6], has a broad range of domain areas included in it. However, it only provides a structure and a set of general concepts upon which domain ontologies

could be constructed. Financial concepts are among these concepts, too. The common terms that we import from this ontology are “FinancialAccount”, “FinancialContract”, “financial asset” and all of their subclasses. The relationships that these terms have with the other classes in our ontology can be seen in Figure 4.

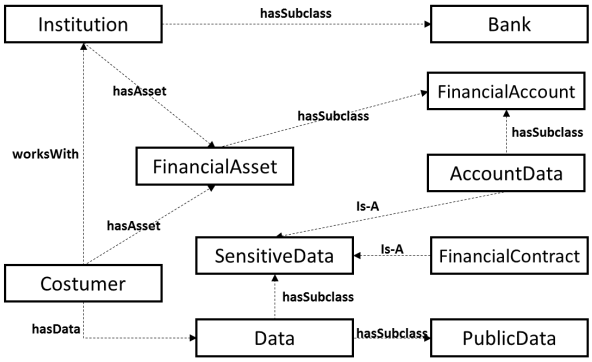

Figure 4: Integration of SUMO ontology classes

Finance ontology [9] is an ontology on financial instruments, involved parties, processes and procedures in securities handling. We are using this ontology to define the financial instruments and involved parties within organizations, so that we do not recreate an already modeled structure and the main concern of our ontology stays as usual banking processes and insider attacks. The common terms that we import from this ontology are “Address”, “Party” and all of the subclasses of “Party” as can be seen in Figure 5.

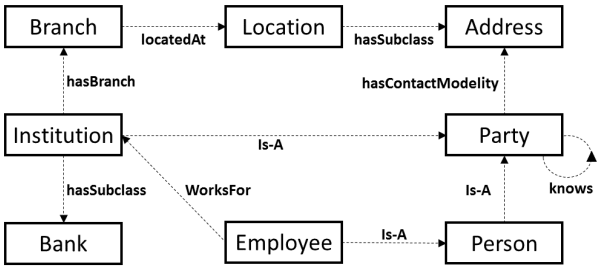

Figure 5: Integration of Finance ontology classes

Therefore, we integrate our ontology with FOAF to base our Person and Organization structure on universally defined terms and we expand these terms. On the other hand, we import classes from SUMO and Finance ontology to use the classes that are already defined in financial domain, so that we do not need to define new classes in the finance domain. The graph of the main components of resulting ontology is shown in Figure 6.

4 Data Sources

Most institutions cannot provide or reveal all the details of insider attacks due to regulatory reasons. Therefore, we validate our initial proof of concept based on the limited data that we are provided with by the collaborating financial institution.

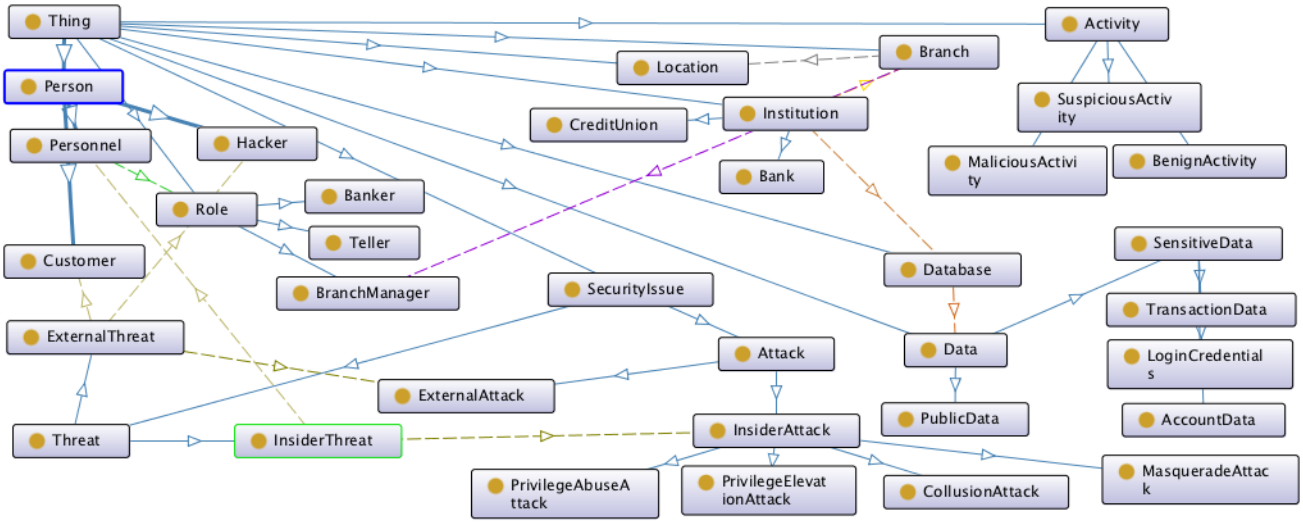

Figure 6: Main components of the ontology

4.1 Databases and log files

We need to understand access patterns for a database in a financial institution database system. This includes a snapshot of the data in the database, as well as query logs including:

- The look up or update query being issued
- (Anonymized) identification of the user or role that accessed the data
- (Anonymized) physical or IP address of the machine issuing the query

What we try to achieve here is to see a view of the daily life of that database. The size and traffic information of the database can be gathered from the log entries. However, we don't have access to the database itself, or its snapshot.

4.2 Organization and permission structure

To have a better understanding of the internal structure of the environment, we create a taxonomy to identify roles and differentiate between different attack types. This taxonomy for the user roles includes the following information:

- User roles
- Access Permissions
- Manages
- Managed by

For instance, what does a bank teller do? What privileges does a teller have on the database? Who manages the teller and who has more privileged access to the system in that branch? Who can change the privileges given to a user? We need to ask these questions for selected types of users. Some of the relationships that are represented in the ontology does not have to present in order to trust the resulting structure. This missing values can be created in time, or left blank.

4.3 Normal behavior

Experimenting on existing logs of a database reveals a lot of information to profile the normal behavior of users. However, taking this data as ground truth means that we accept insider attacks that have not been identified as normal behavior. Hence when we encounter such characteristics, we may classify the intents of those behaviors as benign. As stated before, the ontology structure aims to model the normal world, not the attacks. We use the logs of databases to create normal behavior patterns which then should be reviewed by banking experts.

4.4 Previously identified attacks

We will use attacks identified in the past to create new insider attack scenarios and simulate them on the example databases. Real examples provided by financial institutions will be a guideline preparing these scenarios. The data, log files, and scenario details about some insider attacks that are detected before should be very useful in this phase.

These scenarios can either be used to *create attack models*, or *simulate an attack* to see if the system accurately identifies an insider attack. Although this information would be very beneficial if it existed in order to see how a real attack is represented in the ontology, the experts of the institution we are collaborating expressed that this information is one of the most classified data types.

5 Validation

The proposed insider threat ontology includes the following insider attack types: masquerade attacks, privilege elevation attacks, privilege abuse attacks and collusion attacks. Some attacks may appear in various cases and they may seem very different than each other even if they belong to the same classification, some can even happen unintentionally. According to our study, the classification given above covers all types of insider attacks reported in the literature. If novel attacks are later discovered, additional types can be easily integrated into the ontology framework.

The validation we perform in this paper is made through insider attack scenarios and it aims to test the competency of the ontology in representing any attack that an organization can experience. These scenarios are based on general banking processes and they are not based on real attacks, since we are not provided with that information. To perform this validation, we use description logic (DL) to show how the ontology deals with the information that is provided by the information systems. We choose SROIQ [20] which is considered one of the most expressive DLs and can be considered as a syntactic variant of OWL DL [21]. This DL basically has two kinds of clauses:

- T-Box: it describes the rules of the world as they are described in the knowledge base, and only includes the relationships between classes. For example:

$$employee \sqsubseteq person \quad \text{“employee is a person”}$$

$$customer \sqsubseteq person \quad \text{“customer is a person”}$$

- A-Box: it describes the properties of individuals and their relationships. For example:

$$person(mary) \quad \text{“mary is a person”}$$

$$person(matthew) \quad \text{“matthew is a person”}$$

$$loves(mary, matthew) \quad \text{“mary loves matthew”}$$

We provide a step by step process for each type of attack. For each step, we show

- Plain text: what relevant information we can have from the information systems represented in plain text form.
- Description logic: formal representation of the information we have in the corresponding step.
- Data source: what sources we use while compiling this information.
- Comment: additional explanations if need.

The following part explains the above four attack types and provides an example each.

5.1 Masquerade attacks

In a masquerade attack, the attacker illegitimately assumes the identity of a legitimate user [22]. Before launching an attack like this, the attacker must gather the credentials to access the system. However, the gathering phase of the attack is outside of the scope of this paper, and so we will make the assumption that the attacker has already gained the credentials necessary to access the system. Here, it is clear that the attackers try to hide their identity and make the victim responsible for any action they take. An example scenario is given below and the analysis of the scenario is given in Table 1.

Scenario: Amelia and Ben are tellers and work at the Amherst branch of a bank. Cecile comes to the branch and asks Amelia for her account balance and withdraws some money from her account. Amelia steps out to a meeting with the branch manager and leaves her computer (PC1) logged in to the system thinking that she will be gone for a very short time. Ben takes advantage of her absence, and uses her computer to look up a Dan and Emily's social security numbers and transactions to use them later. He returns back to his seat before Amelia comes back. A few days later, Dan realizes that there is a credit inquiry in his credit report and contacts the bank about the source of the inquiry.

T-Box rules are:

$$Customer \sqsubseteq \exists hasData.CustomerData$$

$$\exists checks.CostumerData \sqsubseteq Employee$$

$$checks \sqsubseteq knows \circ hasData$$

which mean "Every Customer has CustomerData", "Only Employees check customers' data", and "If you check someone's data, you must know that person," respectively.

The other T-Box rules are not needed in the given scenario, hence they are not given.

Table 1: Masquerade attack

#	Plain text	Description logic	Data source	Comment
1	Amelia is a person and a teller	person(Amelia) employee(Amelia) teller(Amelia)	Organizational Structure	
2	Ben is a person and a teller	person(Ben) employee(Ben) teller(Ben)	Organizational Structure	
3	Amelia works at Amherst Branch	branch(Amherst) worksAt(Amelia, Amherst)	Organizational Structure	

4	Ben works at Amherst Branch	worksAt(Ben, Amherst)	Organizational Structure	
5	Amelia works with Ben	worksWith(Amelia, Ben)	Organizational Structure	
6	Ben works with Amelia	worksWith(Ben, Amelia)	Organizational Structure	
7	Amelia uses PC1	equipment(PC1) uses(Amelia, PC1)	Organizational Structure	
8	Ben uses PC2	equipment(PC2) uses(Ben, PC2	Organizational Structure	
9	Cecile checks into Amherst Branch	visits(Cecile, Amherst)	Query Logs	Cecile swipes her card to confirm her identity
10	Cecile is a customer of Amherst Branch	person(Cecile) customer(Cecile) customerOf(Cecile, Amherst)	Database	Cecile's information is retrieved from database
11	Amelia confirms Cecile's identity	knows(Amelia, Cecile)	Query Logs	Interaction creates a relationship between Amelia and Cecile
12	Amelia looks up Cecile's data	account(Cecile's data) hasAccount(Cecile, Cecile's data) checks(Amelia, Cecile's data)	Query Logs	Important point here is that these queries follow the last 2 steps
13	Amelia checks balance for Cecile	checks(Amelia, Cecile's data)	Query Logs	
14	Amelia withdraws money for Cecile	changes(Amelia, Cecile's data)	Query Logs	At the end of this operation, Amelia leaves and Ben uses the computer but our systems don't know it

15	Amelia looks up Dan's data	customer(Dan) account(Dan's data) hasAccount(Dan, Dan's data) checks(Amelia, Dan's data)	Query Logs	Suspicious query should be caught by the system and trigger the ontology, because this is not normal behavior of Amelia.
16	Amelia looks up Emily's data	customer(Emily) account(Emily's data) hasAccount(Emily, Emily's data) checks(Amelia, Emily's data)	Query Logs	Suspicious query should be caught by the system and trigger the ontology, because this is not normal behavior of Amelia. At this point, the events are under investigation.
17	Dan makes a credit inquiry		Query Logs	
18	Dan appeals to the bank		Query Logs	The required actions should take place here according to the result of the investigation

With the T-Box rules given above, the DL requires knows(Amelia, Dan) and knows(Amelia, Emily) which don't exist. The activities in Step 15 and Step 16 violate the rules, hence make the T-Box rules unsatisfiable.

5.2 Privilege elevation attacks

In a privilege elevation (also known as privilege escalation) attack, a user with insufficient permissions accesses the information that only a more privileged user can see. The attackers usually exploit a vulnerability of the system to escalate granted permissions [23], so that they can use these new permissions to access information. An example scenario is given below and the analysis of the scenario is given in Table 2.

Scenario: Ben is a teller at the Amherst branch of a bank. He finds out that after a software update the system allows him to see all of the sensitive information of the bank's customers and thinking that he is allowed to see them, he doesn't notify his superiors. Ben uses these privileges to look up the other teller Dan's sensitive information out of curiosity. However, he does not take things further.

T-Box rules are:

$$Customer \sqsubseteq \exists hasData.CustomerData$$

$$Teller \sqsubseteq \forall checks.CustomerData$$

which mean "Every Customer has CustomerData", and "Every Teller can only check CustomerData," respectively.

The other T-Box rules are not needed in the given scenario, hence they are not given.

Table 2: Privilege elevation attack

#	Plain text	Description logic	Data source	Comment
1	Ben is a person and a teller	person(Ben) employee(Ben) teller(Ben)	Organizational Structure	
2	Ben works at Amherst Branch	branch(Amherst) worksAt(Ben, Amherst)	Organizational Structure	
3	Ben works with Dan	person(Dan) employee(Dan) worksWith(Ben, Dan)	Organizational Structure	
4	Ben uses PC1	equipment(PC1) uses(Ben, PC1)	Organizational Structure	
5	Ben looks up Dan's data	account(Dan's data) hasAccount(Dan, Dan's data) checks(Ben, Dan's data)	Query Logs	Suspicious query should be caught by the system and trigger the ontology, because this is not normal behavior of Ben. Ben looks up an employee's information for the first time. The event will violate the given T-Box rule that specifies a teller can only check customer data.

With the T-Box rules given above, the DL requires Dan to be a customer, not an employee. The activity in Step 5 violates the rules, hence makes the T-Box rules unsatisfiable.

5.3 Privilege abuse attacks

In a privilege abuse attack, the user uses his/her permissions to retrieve information that he/she has no need to know. An example scenario is given below and the analysis of the scenario is given in Table 3.

Scenario: Ben is a branch manager of the Amherst branch at a bank in Buffalo, NY. He also looks up sensitive information of some people from New York, NY. As not to look suspicious, he chooses the customers of a specific branch, keeping in mind that people from the same household tend to travel together and have bank accounts from the same branch.

T-Box rules are:

$$Customer \sqsubseteq \exists hasData.CustomerData$$

$$Customer \sqsubseteq \exists visits.Branch$$

$$Employee \sqsubseteq \exists worksAt.Branch$$

$$hasData \circ checks^{-1} \sqsubseteq visits \circ worksAt^{-1}$$

which mean "Every Customer has CustomerData", "Every Customer visits a Branch", "Every Employee works at a Branch", and "You can only check someone's data, if the owner of the data visits the Branch you work at," respectively.

The other T-Box rules are not needed in the given scenario, hence they are not given.

Table 3: Privilege abuse attack

#	Plain text	Description logic	Data source	Comment
1	Ben is a person and a branch manager	person(Ben) employee(Ben) branchManager(Ben)	Organizational Structure	
2	Ben works at Amherst Branch	branch(Amherst) worksAt(Ben, Amherst)	Organizational Structure	
3	Ben looks up for Dan, a customer of a branch in New York, NY	person(Dan) customer(Dan) account(Dan's data) hasAccount(Dan, Dan's data) checks(Ben, Dan's data)	Query Logs	Suspicious query should be caught by the system and trigger the ontology based system, because this is not normal behavior of Ben. The event will violate the given T-Box rule that specifies the requirement of the customer should visit the branch that the employee works at.

With the T-Box rules given above, the DL requires Dan to visit the branch where Ben works at. The activity in Step 3 violates the rules, hence makes the T-Box rules unsatisfiable.

5.4 Collusion attacks

In a typical collusion attack, there are usually more than one people with different privileges collaborating to access and harvest information [24]. Since this data is usually supposed to include more relations and be more extensive, the impact of these attacks is usually higher. An example scenario is given below and the analysis of the scenario is given in Table 4.

Scenario: Carl is a branch manager and Karen is a secretary at the same branch of a bank. Carl leaks information from the database systems and from internal documents of the bank to a rival company. However, after gathering them, Carl hides the information along with a lot of other information. He orders Karen to collect some specific information and send it to a specific addresses after adding it to the files that he gave her. Karen doesn't know that she is collaborating but she doesn't check out what she is sending.

T-Box rules are:

$$SuspiciousActivity \doteq BenignActivity \sqcup MaliciousActivity$$

$$BenignActivity \sqcup MaliciousActivity \sqsubseteq \bot$$

$$SuspiciousActivity \subseteq \exists mayLeadTo.InsiderThreat$$

$$MaliciousActivity \subseteq \exists leadsTo.InsiderAttack$$

which mean "SuspiciousActivity is either a BenignActivity or MaliciousActivity", "BenignActivity and MaliciousActivity are two disjoint types", "Every SuspiciousActivity may lead to an InsiderThreat", and "Every MaliciousActivity leads to InsiderAttack," respectively.

The other T-Box rules are not needed in the given scenario, hence they are not given. The important point in this scenario to realize is that this time there is no inconsistency in the knowledge base. The system just marks an activity as SuspiciousActivity and it is reported to the security unit.

Table 4: Collusion attack

#	Plain text	Description logic	Data source	Comment
1	Carl is a person and a branch manager	person(Ben) employee(Ben) branchManager(Ben)	Organizational Structure	
2	Karen is a person and a secretary	person(Karen) employee(Karen) secretary(Karen)	Organizational Structure	
3	Carl works at Amherst Branch	branch(Amherst) worksAt(Carl, Amherst)	Organizational Structure	
4	Karen works at Amherst Branch	worksAt(Karen, Amherst)	Organizational Structure	
5	Carl looks up for daily transactions	sensitiveData(dailyTransactions) checks(Carl, dailyTransactions)	Query Logs	This won't be a suspicious query because as a branch manager, Carl is expected to monitor daily transactions

5	Karen looks up for daily branch statistics	sensitiveData(branchStatistics) checks(Karen, branchStatistics)	Query Logs	This won't be a suspicious query because as a secretary, Karen is expected to report on this data. However, we expect our anomaly detection system to identify the harvesting. In that case, the ontology based system is triggered. The system creates a SuspiciousActivity individual in the ontology and reports consecutive access to sensitive data which indicates collusion.

It is important to notice that all attack types has a characteristic violation. Depending on the violations, the ontology can identify the attack type with the rules using the same methodology.

6 Discussion

We have presented an ontology framework focusing on insider attacks in banking domain targeting database systems [3]. We extended our work providing example scenarios on different attack types and justifying how such a system can be useful against insider threats. As indicated before, the prior efforts in insider threats branches to two different directions: psychological aspects and physical aspects of insider threats. Our paper extends the work on building a cyber ontology for insider threats in the financial sector, as it is critical to having capable systems that can identify such threats and developing countermeasures against insider attacks in this domain. The specific contributions of our paper are,

- making use of ontologies as a supporting mechanism to validate the outcome of other anomaly detection systems
- creating a cyber ontology framework for insider threats in the financial sector focusing on relational database management systems
- providing support to the security experts in the investigation process
- ensuring the integration with other knowledge domains to enable data integration.

The literature survey we have conducted shows us that this ontology fills the gap in ontological structuring of insider threat research in the financial sector. The ontologies developed on insider threat research generally focus on defining what an insider threat is and how it takes place, [13, 15]. The work in [15] leads us to experiment on specific domains and use the domain specific knowledge to create a semantic structure while the ontology proposed in [13] shows the relationships between insider threat concepts and insider threat activity. This structure defines the insider threat in financial sector more conclusively. Even if we have collaborated with financial sector experts, we know that there is still a lot to do to expand the capability of our ontology, since we are restricted to gather real data from banking databases. One more way of validating our ontology and consistency of our system is to create a synthetic database. However, to create a synthetic database, we still need to have statistical information of the database tables and columns which is not readily available for us.

The cyber ontology we created has classes from FOAF and SUMO ontologies, which are universally defined, and the terms in them mean the same across all knowledge domains. In this sense, our ontology provides a high level of semantic interoperability. When fully developed, we believe that this integration with other domains and semantic structures approach can prove effective to addressing more factors about insider threats as it could be used by researchers to test and evaluate their detection and mitigation schemes, as well as identifying similar attacks by using previously identified attacks.

The validation scheme we performed is using the same ontology relations and classes. We represent ontology instances in description logic and show that how the ontology captures the differences between different attacks. However, the performance of this system completely depends on another system that profiles user behavior on the database system and catches anomalies in the user behavior depending on the SQL queries that the user sends to the system. Previous studies show that data-centric approach to identify anomalies in user behavior is very effective [19]. However, we are still working on a similar system that does not need direct access to the data to perform analysis.

7 Future Work

Insider attacks have the potential to harm the confidence placed in an organization and acting on suspicions without evidence could cause discontent within the employees. The methodology proposed in this paper aims to exploit the chronology of events and capture the intent behind actions to provide a solid case in case of an attack and protect against false positives. However, to achieve this task, we should create our knowledge base from real working systems to be able to validate the ontology that we constructed. This ontology should be extended based on real data provided by financial institutions. We are working on building collaborations with financial institutions to gather the data required as indicated in Section 4. This data will be used to build on the ontology to improve both scope and capability after the validation of each provided scenario. The current validation scheme uses description logic to show that the current ontology structure is consistent and captures the defined attack types. However, when we have access to real scenarios, we can test the ontology with competency questions to show that it provides a sufficient level of detail and represents the domain well enough.

Moreover, analyzing the risk of an insider attack for a specific type of attack happening may also contribute to the accuracy of the system. This analysis should be performed both from defender's and the attacker's perspective so that the probability of an attack happening can be calculated to be a positive function of the cost to the attacker [25]. The attackers need a preparation time and effort, which means it is directly corraleted to the probability of an attack happening. On the other hand, the attackers try to find the weaknesses in the system and attack a specific part while defenders consider the system as a whole and take security measures according to that. We will consider these factors as well and try to exploit it to create a more effective semantic structure.

Acknowledgments

This material is based in part upon work supported by the National Science Foundation under award number CNS-1409551. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. We would like to thank Thomas Mitchell, Patrick Coonan and Niccoló Meneghetti for their review of related articles and their help in editing.

References

[1] R. H. Anderson and R. Brackney, "Understanding the insider threat." RAND Corporation, March 2005. [Online]. Available: http://www.rand.org/pubs/conf_proceedings/CF196

[2] C. I. T. Center, "2014 U.S. State of Cybercrime Survey," July 2014.

[3] G. Kul and S. Upadhyaya, "A preliminary cyber ontology for insider threats in the financial sector," in *Proc. of the 7th ACM CCS International Workshop on Managing Insider Security Threats (MIST'15), Denver, Colorado, USA*. ACM, October 2015, pp. 75–78.

[4] The insider threat: An introduction to detecting and deterring an insider spy. http://www.fbi.gov/about-us/investigate/counterintelligence/the-insider-threat. [Online; Accessed on December 10, 2015].

[5] I. Niles and A. Pease, "Towards a standard upper ontology," in *Proc. of the 2nd International Conference on Formal Ontology in Information Systems (FOIS '01), Ogunquit, Maine, USA*. ACM, October 2001, pp. 2–9.

[6] A. Pease, I. Niles, and J. Li, "The suggested upper merged ontology: A large ontology for the semantic web and its applications," AAAI, Tech. Rep. WS-02-11, 2002. [Online]. Available: https://www.aaai.org/Papers/Workshops/2002/WS-02-11/WS02-11-011.pdf

[7] J. Golbeck and M. Rothstein, "Linking social networks on the web with FOAF: A semantic web case study." in *Proc. of the 23rd National Conference on Artificial Intelligence (AAAI'08), Chicago, Illinois, USA*. AAAI Press, July 2008.

[8] L. Ding, L. Zhou, T. Finin, and A. Joshi, "How the semantic web is being used: An analysis of FOAF documents," in *Proc. of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05), Big Island, Hawaii, USA*. IEEE, January 2005, pp. 113–122.

[9] E. Vanderlinden. (2015) Finance ontology documentation. http://fadyart.com/ontologies/documentation/finance/index.html. [Online; Accessed on December 10, 2015].

[10] C. L. Huth, D. W. Chadwick, W. R. Claycomb, and I. You, "Guest editorial: A brief overview of data leakage and insider threats," *Information Systems Frontiers*, vol. 15, no. 1, pp. 1–4, March.

[11] J. Szefer, P. Jamkhedkar, D. Perez-Botero, and R. B. Lee, "Cyber defenses for physical attacks and insider threats in cloud computing," in *Proc. of the 9th ACM Symposium on Information, Computer and Communications Security (ASIA CCS '14), Kyoto, Japan*. ACM, June 2014, pp. 519–524.

[12] J. Hunker and C. W. Probst, "Insiders and insider threats – an overview of definitions and mitigation techniques," *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, vol. 2, no. 1, pp. 4–27, March 2011.

[13] D. Mundie, S. Perl, and C. Huth, "Toward an ontology for insider threat research: Varieties of insider threat definitions," in *Proc. of the 3rd Workshop on Socio-Technical Aspects in Security and Trust (STAST'13), New Orleans, Louisiana, USA*. IEEE, June 2013, pp. 26–36.

[14] S. Mathew, S. Upadhyaya, D. Ha, and H. Q. Ngo, "Insider abuse comprehension through capability acquisition graphs," in *Proc. of the 11th International Conference on Information Fusion (FUSION'08), Cologne, Germany*. IEEE, June 2008, pp. 1–8.

[15] D. Costa, M. Collins, S. Perl, M. Albrethsen, G. Silowash, and D. Spooner, "An ontology for insider threat indicators: Development and application," in *Proc. of the 9th Conference on Semantic Technology for Intelligence, Defense, and Security (STIDS'14), Fairfax, Virginia, USA*. CEUR Workshop Proceedings, November 2014, pp. 48–53.

[16] H. Hedden, *The Accidental Taxonomist*. Medford, New Jersey: Information Today, Inc., 2010.

[17] G. Falquet, C. Métral, J. Teller, and C. Tweed, *Ontologies in Urban Development Projects*, ser. Advanced Information and Knowledge Processing. Springer-Verlag London Limited, 2011, vol. 1.

[18] N. F. Noy and D. L. McGuinness. Ontology development 101: A guide to creating your first ontology. http://protege.stanford.edu/publications/ontology_development/ontology101.pdf. [Online; Accessed on December 10, 2015].

[19] S. Mathew, M. Petropoulos, H. Q. Ngo, and S. Upadhyaya, "A data-centric approach to insider attack detection in database systems," in *Proc. of the 13th International Conference on Recent Advances in Intrusion Detection (RAID'10), Ottawa, Ontario, Canada, LNCS*, vol. 6307. Springer Berlin Heidelberg, September 2010, pp. 382–401. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-15512-3_20

[20] I. Horrocks, O. Kutz, and U. Sattler, "The even more irresistible SROIQ," in *Proc. of the 10th International Conference on Principles of Knowledge Representation and Reasoning (KR'06), Lake District, UK*. AAAI Press, June 2006, pp. 57–67.

[21] M. Krotzsch, F. Simancik, and I. Horrocks, "A description logic primer," *arXiv preprint arXiv:1201.4089*, 2012. [Online]. Available: http://arxiv.org/abs/1201.4089

[22] M. B. Salem and S. J. Stolfo, "Modeling user search behavior for masquerade detection," in *Proc. of the 14th International Conference on Recent Advances in Intrusion Detection (RAID'11), Menlo Park, California, USA, LNCS*, vol. 6961. Springer Berlin Heidelberg, 2011, pp. 181–200. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-23644-0_10

[23] L. Davi, A. Dmitrienko, A.-R. Sadeghi, and M. Winandy, "Privilege escalation attacks on Android," in *Information Security: Proc. of the 13th International Conference (ISC'10), Boca Raton, Florida, USA, LNCS*, vol. 6531. Springer Berlin Heidelberg, October 2010, pp. 346–360. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-18178-8_30

[24] A. Sarkar, S. Kohler, S. Riddle, B. Ludaescher, and M. Bishop, "Insider attack identification and prevention using a declarative approach," in *Proc. of the 2014 IEEE Security and Privacy Workshops (SPW'14), San Jose, California, USA*. IEEE, May 2014, pp. 265–276.

[25] A. M. Sanzgiri and shambhu Upadhyaya, "Feasibility of attacks: What is possible in the real world–a framework for threat modeling," in *Proc. of the 2011 International Conference on Security and Management (SAM'11), Las Vegas, Nevada, USA*. CSREA Press, July 2011.

[26] Merriam-Webster.com, "Merriam-webster," http://www.merriam-webster.com/dictionary/term, [Online; Accessed on December 10, 2015].

Author Biography

Gokhan Kul is a PhD student in computer science and engineering at the State University of New York at Buffalo. His research interests include insider threats and knowledge representation. Prior to beginning the PhD program, Gokhan received his M.S. from Middle East Technical University in Turkey and worked as a software engineer there while participating in various projects as an ontology consultant. His current work focuses on threat modeling and insider threat detection under the supervision of Dr. Shambhu Upadhyaya and Dr. Oliver Kennedy.

Shambhu Upadhyaya is a professor of computer science and engineering at the State University of New York at Buffalo where he also directs the Center of Excellence in Information Systems Assurance Research and Education (CEISARE), designated by the National Security Agency. Prior to July 1998, he was a faculty member at the Electrical and Computer Engineering department. His research interests are in broad areas of information assurance, computer security and fault tolerant computing. He has authored or coauthored more than 260 articles in refereed journals and conferences in these areas. His research has been supported by the National Science Foundation, U.S. Air Force Research Laboratory, the U.S. Air Force Office of Scientific Research, DARPA, and National Security Agency. He is a senior member of the IEEE.

A Taxonomy definitions

The generalized terms in the taxonomy we created for insider threats is given below:

Location: A position or site occupied or available for occupancy or marked by some distinguishing feature [26]

Institution: An established organization or corporation especially of a public character [26]

Bank: An establishment for the custody, loan, exchange, or issue of money, for the extension of credit, and for facilitating the transmission of funds [26]

Credit Union: A cooperative association that makes small loans to its members at low interest rates and offers other banking services [26]

Branch: A local office at a specific location of an institution

Person: One (as a human being, a partnership, or a corporation) that is recognized by law as the subject of rights and duties [26]

Customer: One that purchases a commodity or service [26]

Hacker: A person who illegally gains access to and sometimes tampers with information in a computer system [26]

Personnel: The people who work for a particular company or organization [26]

Role: A function or part performed especially in a particular operation or process [26]

Teller: A person who works in a bank and whose job is to receive money from customers and pay out money to customers [26]

Banker: A person that engages in the business of banking [26]

Branch Manager: A person that is responsible for managing a branch of an institution and the personnel working at that branch

Data: Facts or information used usually to calculate, analyze, or plan something [26]

Public Data: Data that can be accessed by anyone who is interested. The access and usage rights may vary and can be accessed with various ways

Sensitive Data: Data that calling for care and caution which can usually cause problems in case that someone else uses it

Account Data: Data that belongs to personal or business accounts which includes but not limited to name, address, account number etc. [26]

Login Credentials: Data that belongs to personal or business accounts which includes login usernames and passwords, security questions and answers

Transaction Data: Data of "a communicative action or activity involving two parties or things that reciprocally affect or influence each other" [26]

Database: A usually large collection of data organized especially for rapid search and retrieval (as by a computer) [26]

Attack: To set upon or work against forcefully [26]

Threat: Someone or something that could cause trouble, harm, etc. [26]

Activity: Something that is done as work or for a particular purpose [26]

SuspiciousActivity: An activity that tends to arouse suspicions about what the user intended

BenignActivity: A completely normal behavior or an activity that looks suspicious but actually normal

MaliciousActivity: An activity that has an intent to attack the system

Security Issue: A matter or event of threat or attack

External Threat: A threat that is posed by someone or something that is not from the personnel of an institution

Insider Threat: Malevolent (or possibly inadvertent) actions by an already trusted person with access to sensitive information and information systems [1]