AZALIA: an A to Z Assessment of the Likelihood of Insider Attack

Matt Bishop
University of California Davis
Davis, CA
Email: bishop@cs.ucdavis.edu

Carrie Gates
CA Labs
New York, NY
Email: carrie.gates@ca.com

Deb Frincke and Frank L. Greitzer
Pacific Northwest National Laboratory
Richland, WA
Email: deb.frincke@pnl.gov, frank.greitzer@pnl.gov

***Abstract*—The insider threat problem is increasing, both in terms of the number of incidents and their financial impact. To date, solutions have been developed to detect specific instances of insider attacks (*e.g.*, fraud detection) and therefore use very limited information for input. In this paper we describe an architecture for an enterprise-level solution that incorporates data from multiple sources. The unique aspects of this solution include the prioritization of resources based on the business value of the protected assets, and the use of psychological indicators and language affectation analysis to predict insider attacks. The goal of this architecture is not to detect that insider abuse has occurred, but rather to determine how to prioritize monitoring activities, giving priority to scrutinizing those whose background includes access to key combinations of assets as well as those psychological/other factors that have in the past been associated with malicious insiders.**

I. Introduction

Recent surveys indicate that the "financial impact and operating losses due to insider intrusions are increasing" [1]. Within the government, insider abuse by those with access to sensitive or classified material can be particularly damaging. Further, the detection of such abuse is becoming more difficult due to other influences, such as out-sourcing, social networking and mobile computing. Traditional notions of an insider have focused on defining an insider based on employment; however, this notion has two limitations. First, it does not take into account access via outsourcing, contractors, etc., where proprietary information has (legally) left the organization. Second, it treats masqueradors (those people who have illegally gained access to an insider s account and is now abusing that access) and traitors (traditional insiders) as separate threats, even though they are using the same accounts and accessing the same resources. We therefore use the definition of insider as described by Bishop and Gates [2], where an insider is defined based on *access*.

Current solutions for addressing insider threat are limited, specifically:

1) No enterprise-level solution has been developed, but rather only point solutions. That is, current solutions use a restricted set of input data and focus on particular types of insider abuse.
2) Current solutions treat all threats equally. The result is that organizations do not know how to prioritize their resources in order to address their greatest threats and protect their most important assets.
3) Current solutions focus on identification, not prediction, of threats. The result is that the damage has often already occurred before the abuser has been identified.
4) Current solutions focus strictly on transactional analysis. The result is that the psychological indicators that might be present in language are missed.

This paper will focus on a key aspect of our enterprise-wide architecture: a risk assessment based on predictions of the likelihood that a specific user poses an increased risk of behaving in a manner that is inconsistent with the organizations stated goals and interests. We will present a high-level architectural description for an enterprise-level insider threat product.

We then focus on the predictive capabilities that we are investigating, which are predicated on psychological analyses. Specifically, we discuss what data needs to be collected in order to recognize possible changes in the psychological state of a particular individual, as well as the feasibility of collecting such information and analyzing it through automated means.

It is expected that indications of psychological changes will be apparent in free-form text, such as email. We therefore discuss approaches to analyzing content to determine if we can detect indications of psychological state from it and, more importantly, if we can detect changes in psychological state, such as a person becoming increasingly disenfranchised with their organization.

After presenting the generic architecture, we focus specifically on our prioritization model. Previous work by the authors has focused on an access control-like model that prioritizes resources based on business value to the organization. Users are then grouped according to their access to resources and ordered based on the business value of what they can access. We will describe our model in greater detail, as well as discussing how we intend to extend this model to take into account the psychological indicators of threat associated with each user.

Previous work on insider threat detection has focused on very specific problem spaces. In this paper we will provide an enterprise-level architecture with a focus on predicting users that are at greatest risk of becoming a threat, balanced against the access they have to resources that are of the greatest value to an organization. We present our architecture in Section II.

In Section III we provide details on previous work on the analysis of psychological indicators that were present in people convicted of insider abuse, and describe the inputs we can use for performing an automated analysis of insider psychological state. We then go on in Section IV to briefly discuss language affectation analysis, and how this might be applied as another input to our overall architecture. In Section V we provide previous work on a model for prioritizing monitoring and analysis capabilities based on the associated business value of the resouces being monitored. We then descibe how this model can be extended to use the psychological indicators in order to prioritize and predict threats. We describe an example use case for the overall system in Section VI, followed by providing some concluding comments in Section VII.

II. Enterprise Architecture

A key shortcoming of current insider threat approaches is that they focus on point solutions that employ algorithms requiring specific input and having specific goals. A generic approach to the insider threat problem has not been developed.

We take an enterprise-wide approach to solving the insider threat problem, with the goal of predicting threats based on associated risks rather than detecting attacks after the event has occurred. This approach allows management to both focus resources on monitoring those assets of greatest value and those insiders that pose the greatest threat, as well as providing them with the possibility of addressing insider concerns in order to reduce the threat they pose. Figure 1 presents a generalized architecture for our approach using the C3A format [3]. We note the presence of five key modules (which we describe in more detail below), however this paper focuses primarily on only two of these modules (the threat management and threat indicators modules).

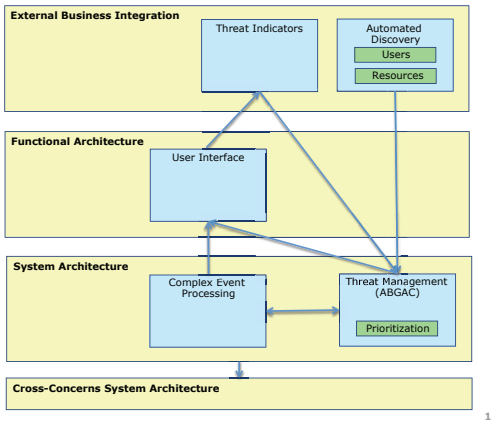

Fig. 1. High-level reference architecture for insider threat

Given that we expect to develop algorithms that are appropriate for an enterprise environment, we start with the automated discovery module. The goal of this module is to determine the users and resources that exist in the enterprise. This information can be collected from a variety of sources, such as from existing identity and access control systems, or deduced from an analysis of log files (or some combination of these two processes). The threat indicators module provides information about users and systems that indicate potential risk. For example, this module might take advantage of known blacklists, or the security personnel might manually input specific users or resources that pose increased threat. This module uses analysis of psychological indicators (described in more detail in Section III) and language affectation (described in more detail in Section IV) in order to provide additional threat information. These two modules are part of the external business integration layer because they interface with other systems within the organization.

These two modules feed the threat management (ABGAC) module in the system architecture layer. The threat management module is based on the ABGAC (Attribute-Based Group Access Control) model originally developed by Bishop and Gates [2] and described in more detail in Section V. In summary, this module develops a matrix of users versus resources where each element of the matrix records the access that user has to that resource. The users, resources (assets) and access come from the automated discovery module. In the traditional ABGAC model, the assets are prioritized based on their business value (such as being based on the strategic importance of information in a database, rather than the value of the hardware and software). This provides an ordered list of assets. Users are grouped based on their access to these assets. The security officer can then focus resources on those users with access to the assets of greatest business value in a more standard risk management practice.

In this paper we describe an extension to the ABGAC model by identifying risk factors that can be assigned to users, resulting in a weighting associated with each user that indicates their propensity towards acting in a manner that is inconsistent with the organizations security objectives. This information can then be combined with the business value of resources to determine how best to address the threat (such as through greater logging on the individual to determine if they are posing a threat, to more direct measures such as approaching the individual).

The threat management module interacts with a complex event processing module that is responsible for event collection and real-time analysis, as well as with event logging and off-line analysis. It provides information back to the threat management module so that it can potentially adjust its risk assessment, and takes the risk assessment information from this module in order to determine if it should modify its event processing and logging procedures in a self-adaptive manner.

The complex event processing and threat management modules belong to the systems architecture layer because they are the main underlying components of the entire system. This layer interfaces with the functional architecture layer by having its two modules provide input to the user interface module. The user interface module provides output to the security officer

indicating if there is an increase in the threat posed by any individual or any activity that should be investigated further. This module also takes input from the security officer that might influence the threat indicators module.

Finally, the cross-concerns system architecture layer provides the standard underlying functionality of any generic system, such as a single sign-on capability or data stores (for example).

While this architecture describes a generic, enterprise-level system, we focus in this paper on two specific modules: threat indicators and threat management. Sections III and IV describe some of the threat indicators that we might use, along with the algorithms for determining how to measure these threats. Section V describes the ABGAC model from the threat management module, along with how it might be extended to take into account the new threat indicators.

III. Threat Prediction and Psychological Indicators

While our overarching definition of insider includes both the traditional notion of individuals who currently or at one time had legitimate authority to act, as well as those who may have gained such capability illicitly, psychological indicators are most apt to be useful in the case where the insider is an individual recognized and trackable by the organization in some capacity [4] (although a masquerader or usurper might be identified through a change in such indicators). Insiders who are trusted employees represent an especially insidious threat to organizations, since they are given access to information that could compromise the organization if it falls into the wrong hands. There has been much research into the psychology and motivation of insiders, but the hard fact is that it is very difficult to predict who will commit security fraud [5]. As noted by Shaw and Fischer [6], "Eighty percent of the threats could have been prevented by timely and effective action to address the anger, pain, anxiety, or psychological impairment of perpetrators, who exhibited signs of vulnerability or risk well in advance of the crime of abuse." Clearly, prediction of insider adverse actions is a difficult challenge, but the potential payoff in avoiding asset and information loss of major consequence serves to justify further research and development of predictive mechanisms. In our work, we seek to combine predictive and detection approaches so as to provide earlier warnings, as well as to reduce the number of false positives that need to be considered. This also is important for insiders who are employees, partly due to the potential negative effects a false suspicion can bring about.

A review of existing insider threat detection/mitigation tools and methods and assessment of current practice is provided by Greitzer *et al.* [7], summarized as follows:

- **Level of Analysis.** Most insider threat tools are focused on a "gross behavioral" level of detection that essentially constrains them to collect forensic data for use in prosecuting the insider who performed the malicious act. Defining possible observable precursors using behavioral cyber indicators is essential for recognizing and predicting inappropriate, suspicious, or disallowed behavior, so that any issues can be addressed before abuse occurs.
- **Methodologies Employed**. Signature based tools have been effective in a reactive role but not very successful as a preemptive protection mechanism, requiring as they do detailed knowledge about a specific behavior or attack. Best practices include deploying systems that use behavioral based analysis as well as signature based analysis of information flows.
- **Interoperability/Integration**. Deployed tools are usually independent products.
- **False Positives**. The reduction of false positives may be better addressed by analyzing behavioral indicators. While there are few opportunities to observe the actual (culminating) malicious acts, there are clearly more opportunities to observe smaller precursor events (indicators) that lead to the eventual exploit.
- **Adaptation**. Most of the products and systems on the market use static rules and thresholds that do not adequately address the dynamic nature of a large enterprise network environment, enabling malicious traffic to flow "under the radar" and out of the purview of the protection and detection systems.
- **Type of Data**. Automated monitoring/detection approaches are uniquely suited to handle the large volumes of data to infer whether the underlying process is benign or malicious. However, factors that are implicated in malicious insider motivation and behavior can provide important contextual information to incorporate into risk analysis and intervention decisions – social/organizational factors that are not based in cyber data but are considered precursors or warning signs. An effective predictive approach to insider threat mitigation must take psychological and organizational factors into account, and while some of these factors may manifest themselves in computer behavior, others do not.

Our framework comprises a knowledge base of what might be called the semantics of insider behavior and characteristics, certain organizational factors that impact motivation and behavior, a large collection of what we shall refer to as "indicators" that reflect intentions and actions, and what we shall refer to as heuristic models of insider behavior. This knowledge base informs the threat management module, and is updated through components handling functions such as data collection, data fusion, analysis, and decision-making (which leads to actions). We assume a process in which data are monitored, collected and analyzed to infer "indicators" or precursors of possible malicious behavior [7]. Some indicators carry more weight than others, while the presence of multiple indicators may create a picture that can only be seen when they are joined. To help articulate the basic modeling concepts, it is useful to think of the process as a multi-layered analysis/inference process that progresses from Data to Observations to Indicators to Behaviors, as depicted in Figure 2.

Fig. 2. Model-Based Predictive Classification Concept: Incoming data processed to infer observations; observations processed to infer indicators; indicators assessed to gauge threat, after Greitzer *et al.* [7].

Data are processed to infer observations. Examples of cyber data include activity of an employee s network account, such as Web traffic or outgoing/incoming data through the firewall. An algorithm may calculate the amount of such Web traffic over time and compute trends in amount downloaded or uploaded, or ratio of upload to download and track changes in such variables over time. Resultant figures may be considered observations such as "amount of Internet Web surfing" or "amount of downloads." Other examples of observations derived from data may include time at work or hours worked (derived from timecard records). Social/organizational data, such as certain HR or performance review records or events, could also be processed to derive observations.

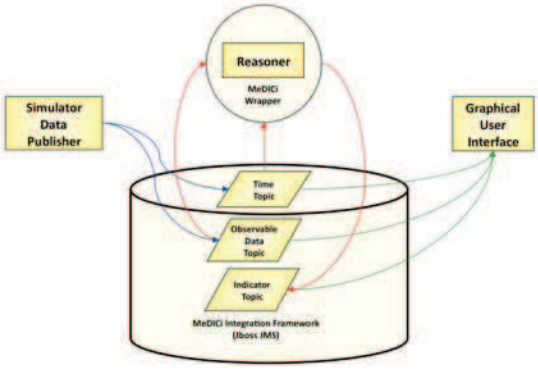

Fig. 3. Architecture implemented for Insider Threat predictive model [4].

Observations are processed to infer indicators. For example, observations relating to Web traffic may be analyzed to derive indicators that represent "possible suspicious" activities such as "increased downloads above normal" or "unusual/late hours worked." Other examples of reasoning about observations to infer indicators are "excessive attempts to access privileged data base," "presence of automated scripts" or "use of personal email account." On the psychosocial side, we may identify an indicator such as "anger in the workplace" based on data and observations such as entries in a HR database relating to arguments with supervisors; or "disgruntled employee" that represents a staff member who exhibits various manifestations/indicators of anger in the workplace.

Indicators are actual malicious acts or precursors to malicious activity; they may be observed directly or inferred from observations.

Indicators are examined to infer behaviors. Behaviors are sequences of activities for achieving some specific purpose

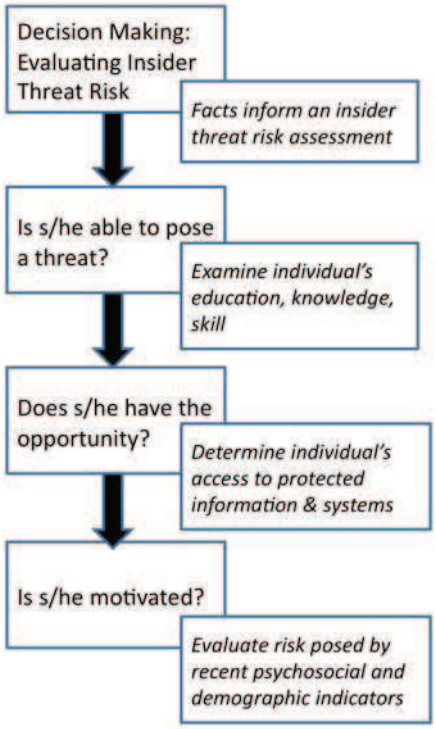

Fig. 4. Assessing ability, opportunity and motivation is a primary decision making task underlying the threat analysis [4].

that can be characterized as malicious or benign. We are most interested in suspicious behaviors that are consistent with established patterns or profiles exhibited in insider exploits. Examples of malicious behaviors are combinations/sequences of behaviors, indicators, etc., such as manifestations of abuse (like attempting to circumvent policy by accessing data without privilege). It is often the case that the elements (observations, data) making up normal work activity are much the same as those that comprise malicious activity. In many cases, then, it is the combination of such observations that lead to increased suspicion that behaviors reflect malicious intent. Only the most blatant acts (*e.g.*, downloading a classified document to a thumb drive when there is an expressed policy forbidding this) can be recognized without a more sophisticated level of analysis involving inference/classification such as we propose.

Our conceptual model employs a hybrid approach that is based on pattern recognition processes, but not merely dependent on identifying discrepancies from "normal" behavior as the primary means of threat analysis. Rather, the knowledge base is populated with scenarios or behavioral templates that reflect possible malicious exploits. While deviation from norms is considered as part of the analysis, so is conformance with prototypical exploits and behaviors that have been identified, through extensive research, with malicious intentions and actions. The challenge is to conduct model-based reasoning on the recognized patterns, at a higher semantic level of concepts/constructs, rather than applying fixed recognition

processes to fixed signatures[1]. The envisioned functionality is accomplished through sophisticated reasoning and adaptive components. Figure 3 illustrates the current architecture of the Insider Threat predictive modeling system, which is provided within a Service-Oriented Architecture (see Greitzer *et al.* [4] for more details).

Cyber Data. The tools that we have reviewed in previous work collectively provide hundreds if not thousands of data elements to include, for example: registry entries, Intrusion Detection System events, firewall logs, DNS logs/Internet sites accessed, network print logs, Web server logs, access to account, email headers, instant messaging, and proximity card data. This data may be all that is available in cases where the "insider" is not part of the organizational structure.

Social/Organizational Data. Numerous studies have been carried out to identify the psychological profiles that are consistent with or possibly predict insider espionage [8], [9], [10][2] that have revealed behaviors, motivations, personality characteristics, and mindsets associated with this criminal behavior. From this body of research, we have synthesized a set of "warning signs" – psychosocial/behavioral indicators – that might be observed prior to an employee actually committing an insider attack [11]. Two basic sources of internal employee data common to most large organizations are human resources data and security data.[3]

A Human Resources Information System, or HRIS, collects, maintains, and reports employee demographics and other data. Larger organizations typically have multiple systems, databases, and processes to manage employee data in addition to the main HRIS. Data collected often includes : national origin/visa status; results of background investigations; personal references; requests for personal, medical, and other leaves; education level; life events; attendance records; performance evaluations; legal issues (*i.e.*, garnished wages); disciplinary issues; complaints by or against the employee; employment applications (includes background check, references, education and work history). Security data typically collected by organizations that handle classified and/or sensitive materials include incidents and indicators in both personnel and electronic security.

IV. Language Affectation Analysis

Advances in social media and collaborative technologies have transformed the workplace. Employees now communicate through a variety of means; instant messaging, email, and even blog postings. Because these forms of communication are prolific and often casual, they provide a better view of employee attitude and satisfaction than more traditional means of workplace communications [12]. There is a human aspect to understanding insider threat; for one to be a threat they need to have a means and motivation. We propose that the investigation of social media, such as email and instant messaging, can help determine which insiders have the means or motivation, which will thus serve as indicators in the overall architecture.

There are two major technological aspects to analyzing social media for motivational content. The first is putting together a suite of tools that can ingest various types of social media and isolate the text and other features that can be used for analysis. There exist a variety of methods to capture chat and email content, as well as blogs [13], [14], [15]. We propose to build on these technologies, using the results as input to the threat indicators module in our architecture.

The second aspect of analyzing these media for insider threat is identifying the linguistic indicators of motivation. While we plan on using a variety of natural language processing tools (e.g., Topic Identification, key word searches, and Named Entity Extraction) to isolate content in these media, the crux of our analysis will focus on identifying markers of attitude and personality in the text. Recent advances in text analysis have led to finer-grained semantic classification, which enables the automatic exploration of subtle areas of meaning. One area that has received a lot of attention is automatic sentiment analysis – the task of classifying documents, or chunks of text, into emotive categories, such as positive or negative. Sentiment analysis is generally used for tracking people s attitudes about particular individuals or items. For example, corporations use sentiment analysis to determine employee attitude and customer satisfaction with their products. We aim to use it as an indicator for motivation of insider threat.

Our approach to sentiment analysis is based on identifying words in the data that convey sentiment or attitude through the use of a lexical look-up method. However, determining motivation for potential insider threat requires much more detailed analysis after this step. We need to analyze sentiment by communication variables (whether the email, for example, is to one s boss or friend), topic (to identify when a person is much more negative about a topic than their peers, for example), and over time to see if attitude changes over time. The focus of our sentiment analysis research is on identifying automatic methods to analyze these aspects of social media so that they can be used as reliable indicators of behavior.

V. ABGAC Model for Threat Prioritization

In our system we employ the Attribute-Based Group Access Control (ABGAC) model originally developed by Bishop and Gates [2] and defined in more detail in a later paper by Bishop *et al* [16]. This model is a generalization of role-based access control (RBAC) [17], [18], except that it uses groups of users

[1]The problem is essentially identical to the early discussion of human cognitive/perceptual systems that argue against a pure template-recognition model that would have to store an essentially infinite number of variations of a concept (such as for "chair") rather than its attributes and functional features or behaviors (one can recognize a chair even if that particular instance has never been seen before).

[2]Project Slammer is a CIA-sponsored study of Americans convicted of espionage against the United States. A declassified interim report dated 14 April 1990 is available at: http://antipolygraph.org/documents/slammer-12-04-1990.shtml and http://antipolygraph.org/documents/slammer-12-04-1990.pdf.

[3]More details about sources of these data and policy, ethical, and legal issues that must be considered in using (and not using) such data are discussed in [10].

based on similar attributes or access rights, rather than by roles (which often include exceptions).

The goal of the ABGAC model is to define groups of resources and, for each group, to define a set of users who have access to that group. The resource groups are ordered by business value, thus creating a corresponding ordered list of user groups. The user groups that have access to the resource groups of greatest value pose the greatest risk to the organization.

For our purposes, the attributes of interest are descriptions of the protection domain of entities. Here, we mean "protection domain" in its broadest sense, not simply a technological listing of rights from capability lists (C-Lists) or access control lists (ACLs). So, the protection domain can include access rights to resources (systems, printers), documents, buildings, and generally any other object to which a user can have access. The protection domain can also include procedural access rights such as physical presence, or the ability to block access. For the purpose of this proposal, however, we focus our efforts on any access that has an associated cyber log file. (For example, if access to a particular room is controlled via a proximity card or biometric, and any access made is logged, then we can use this information. However, if only a standard key is used, or the logging performed is only manually recorded in a physical log book, then we do not use this information, even though the model itself can incorporate it.)

From this, we define a *resource pair* as a pair consisting of a resource (entity) and an access mode describing one way in which that entity can be accessed. For example, a pair might be (*printer, write*), which indicates the ability to write to a printer.

We then define a *resource domain* as a set of resource pairs. This describes a domain similar to the usual notion of protection domain, but includes physical and procedural access as well as cyber access. It is oriented towards the resource (object), not the process (subject).

Once defined, the resource domains need to be ordered. This enables the organization to analyze the cost of restricting access to a particular resource and the benefit of restricting access to that resource, and balance the two. The ordering might be total, such as a linear ordering, or partial, using a vector of measurements taken over different axes. The value of resource domains should be based on its business value, not on its physical value. Thus a laptop that contains the email correspondence for the CEO of an organization is more valuable than the same laptop containing only computer games.

Once ordered, the resource domains can be combined into groups (containing a *contiguous* set of access control settings so that the order is maintained), where the group indicates the threat level a particular set of attributes represents. We therefore define an *rd-group* as a set of resource domains. Different resource domains may be related for the purposes of analysis, although singleton rd-groups (groups consisting of exactly one resource domain) will also prove useful.

Each rd-group induces a set of subjects that have all elements of the rd-group as subsets of the subject s protection domain. We therefore define the *user group* as the set of all subjects whose protection domains are a (possibly improper) superset of the associated rd-group. Note that user groups are created based on the protection domains of the associated users rather than on the job functions of the associated users (as in a role-based system). **The users with access to the rd-groups with the highest value therefore represent those users who pose the greatest risk for insider threat.** There is a natural ordering of user groups based on set containment.

VI. Example Scenario

We examine a medium-sized business as the organization in our example scenario. In this case, the assets identified (at a high level) as having the greatest value to the business were the organization s financial information and the CxO s email (.e., CEO, CFO, etc.). This information is stored on (or accessible via) certain desktops, laptops, backups and printers with read access and physical access, and potentially write access. Given this set of access rights and items, an example resource pair would be (*desktop, read*). An example resource domain might be the CxO s email, which would consist of all of the associated resource pairs (*e.g.*, (*laptop, read*), (*printer, physical access*), etc.). An example rd-group would be both the CxO s email and the financial information.

Beyond these two resource domains (CxO s email and financial inforamtion), there are also other domains, whose value to the business is not as great. For example, these domains might be the IT trouble ticket system, marketing literature and the expense reporting system. Each of these three domains would consist of appropriate resource pairs that define their content. The five resource domains can be ordered along the x-axis based on their value to the business.

On the y-axis are the users of the system. In this example, we have four users: Alice, Bob, Charlie and Eve. Both Alice and Bob are senior system administrators, while Charlie and Eve are junior system administrators. As defined by their roles, Alice and Bob have access to all five resource domains, while Charlie and Eve only have access to the three resource domains of lesser business value. This is consistent with standard role-based access control (RBAC) systems, with the exception that the x-axis consists of higher-level resource domains rather than resources.

However, suppose the CEO does not trust Bob, and so, even though he has the role of senior system administrator, he does not have access to the two high-value resource domains. At the same time, Charlie is being primed for a promotion to a senior position, and so does have access to these two high-value resource domains. Thus our user groups consist of Bob and Eve as one group and Charlie and Alice as the other group. Note that we define user groups based on access permissions, as opposed to defining access permissions based on roles; thus we deviate from RBAC in a significant manner. The matrix that has been defined is demonstrated in Figure 5.

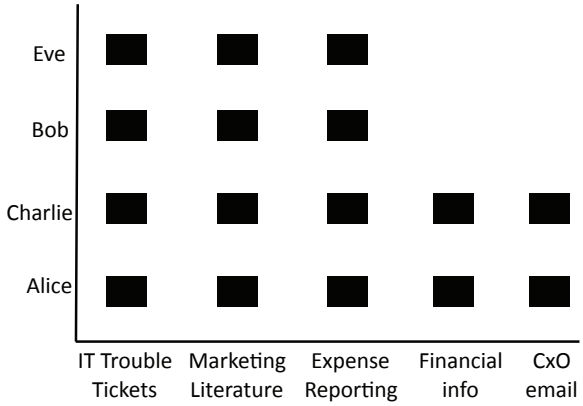

Fig. 5. ABGAC Matrix

Given that the financial information and the CxO email are the assets of greatest business value, this indicates that Alice and Charlie are the insiders who pose the greatest risk to the company. Compounding this at a technical level is the fact that it is not Alice who has access to these assets, but rather the account *alice*. While no one other than Alice should have access to this account, the reality is that other people may also gain access to it. For example, should Alice be working from home and leave her computer unlocked, her spouse might then have access to the account.

This example shows how we can create groups of user accounts (insiders) ordered by the risk they pose to the business based on the value of the assets to which they have access. The users and assets can be discovered automatically via the automated discovery module (*e.g.*, by analysing log files, identity and access management information, etc.). An approach, however, to automatically determining the business value of the information remains as an open research question. Regardless, this input is then provided to the threat management module, which creates the matrix described above, and thus generates an ordered list of user groups.

The threat management module also takes as input threat indicators. This information is gathered from psychological indicators (as described in Section III) and from an analysis of the language used in emails, blogs, etc. For each user a threat value is provided by the threat indicators module, where the threat value is derived from the results of the indicators. The greater the threat value, the more indicators that a user exhibits.

The threat value is combined with the user group information to determine a new ordering of users based on threat as well as risk. The advantage of having users ranked in such a fashion is that security personnel can prioritize their resources, such as monitoring and analysis, focusing on protecting those assets of greatest value from those users posing the greatest threat.

To continue with our example, suppose that Bob is angry that, while he is a senior system administrator, he is not trusted with having access to all of the assets of the company. He is further angered by the fact that Charlie, who is junior to him, has greater access than he does. Bob s disillusionment is evidenced by his often being late for work (which is noted by the times at which he logs into the systems in the morning) and by his not putting in as many hours after work (which is noted by a decrease in his after hours remote access). Further, Bob has been exchanging emails with Eve, where he often vents about his position. Language affectation analysis of his email indicates this change in mental attitude. Based on these indicators, which are significantly different and more extreme than similar activities by the other administrators, Bob is assigned a high threat value, while the other administrators all have a low threat value.

Based purely on risk, the ordering of users (from highest risk to lowest) is (Alice, Charlie), (Bob, Eve). However, by adding the threat value as an additional input to the ordering algorithm, thus creating an order based on threat, the order changes to Bob, (Alice, Charlie), Eve. Thus Bob poses the greatest threat to the company. This can result in several different actions. One action is that security personnel increase monitoring of Bob s actions. Another action is that the human resources department might talk with Bob to try to ease his anger with the company and address the underlying issues.

VII. Conclusions

We presented a enterprise-level architecture that describes a system for prioritizing resources when protecting against insider threats. More specifically, our architecture generates a matrix of assets and the users who have access to those assets. The assets are prioritized based on their value to the business. Users are then grouped based on their access to similar resources (rather than the role they have within an organization), and these groups are ordered based on the value of the assets to which they have access. Thus the users who have access to the resources of greatest value pose the greatest risk to an organization.

This initial ordering of users is combined with threat indicators to form an ordering based on threat. Indicators can consist of psychological indicators gleaned from log files, human resources records, etc., that indicate user disillusionment with their organization. Additional information can be gathered by performing an analysis of language affectation in user emails, instant messages, blog entries, etc. These indicators are combined to generate a threat rating associated with each user, which is then in turn combined with the risk ordering. Thus this architecture provides input to security personnel on who is most likely to commit insider abuse and the degree of damage this might cause.

Unlike previous solutions to the insider threat problem, our approach has the following advantages:

1) We provide an enterprise-level solution, rather than a point solution designed to address specific cases of insider abuse.

2) We prioritize insiders based on the threat they pose to the organization.
3) We use psychological indicators, including language affectation analysis, rather than focusing exclusively on transactional analysis.
4) We predict threats, providing security personnel with the time to address the threat, rather than detecting abuse after it has occurred.

These four advantages outline the key contributions of our work. Unlike other approaches, which focus on detecting insider attacks, we provide an architecture that focuses on providing a list of users ordered by the threat they present. This allows security personnel to prioritize their resources to protect those assets of greatest value and focus on monitoring those users posing the greatest risk. Human resources personnel can also be called upon to address those users posing the greatest threat with the aim of reducing that threat by addressing the underlying dissatisfaction those users have with the organization.

REFERENCES

[1] F. Greitzer, A. Moore, D. Cappelli, D. Andrews, L. Carroll, and T. Hull, "Combating the insider cyber threat," *IEEE Security & Privacy*, vol. 6, no. 1, pp. 61–64, 2008.

[2] M. Bishop and C. Gates, "Defining the insider threat," in *Proceedings of the 2008 Cyber Security and Information Infrastructure Research Workshop*, 2008.

[3] E. Hadar and G. Silberman, "Agile architecture methodology: Long term strategy interleaved with short term tactics," in *Proceedings of the International Conference on Object Oriented Programming, Systems, Languages and Applications (OOPSLA)*, 2008.

[4] F. Greitzer, P. Paulson, L. Kangas, T. Edgar, M. Zabriskie, L. Franklin, and D. Frincke, "Predictive modelling for insider threat mitigation," Pacific Northwest National Laboratory, Richland, WA, Tech. Rep. PNNL Technical Report PNNL-60737, June 2008.

[5] L. Kramer, R. H. Jr., and K. Crawford, "Technological, social, and economic trends that are increasing u.s. vulnerability to insider espionage," Personnel Security Research Center (PERSEREC), Tech. Rep. Technical Report 05-10, May 2005.

[6] E. Shaw and L. Fisher, "Ten tales of betrayal: The threat to corporate infrastructures by information technology insiders: Report 1 - overview and general observation," Personnel Security Research Center (PERSEREC), Monterey, CA, Tech. Rep. Technical Report 05-04, April 2005.

[7] F. Greitzer, L. Kangas, T. Edgar, A. Brothers, and P. Paulson, "Predictive adaptive classification model for analysis and notification: Insider threat (pacman:it)," Pacific Northwest National Laboratory, Richland, WA, Tech. Rep. PNNL Technical Report PNNL-16713, July 2007.

[8] D. Parker, *Fighting Computer Crime: A New Framework for Protecting Information*. New York, NY: John Wiley & Sons, Inc, 1998.

[9] M. Gelles, "Exploring the mind of the spy," http://www.dss.mil/search-dir/training/csg/security/Treason/Mind.htm, Texas A&M niversity Research Foundation, 2005, last Visited: 18 April 2009.

[10] Redacted, "Project SLAMMER interim report," http://antipolygraph.org/documents/slammer-12-04-1990.pdf, April 1990, last Visited: 18 April 2009.

[11] F. Greitzer, P. Paulson, L. Kangas, L. Franklin, T. Edgar, and D. Frincke, "Predictive modelling for insider threat mitigation," Pacific Northwest National Laboratory, Richland, WA, Tech. Rep. PNNL Technical Report PNNL-65204, March 2009.

[12] A. Cowell and M. L. Gregory, "nvestigations in collaborative multi-party discourse," in *Proceedings of the International Conference on Weblogs and Social Media*, March 2007.

[13] M. L. Gregory, D. V. Love, and S. Rose, "Analyzing threaded dialogue to aid with network security protection," November 2006, demo presented at *SuperComputing 2006*.

[14] M. L. Gregory, D. Payne, D. McColgin, D. Love, and N. Cramer, "Visual analysis of weblog content," in *International Conference on Weblogs and Social Media*, March 2007.

[15] A. J. Cowell, M. L. Gregory, E. J. Marshall, and L. McGrath, "Knowledge encapsulation framework for collaborative social modeling," in *AAAI Spring Symposium on Techno-social Predictive Analytics*, 2008, in submission.

[16] M. Bishop, S. Engle, S. Peisert, S. Whalen, and C. Gates, "We have met the enemy and he is us," in *Proceedings of the 2008 New Security Paradigms Workshop*, September 2008.

[17] D. Ferraiolo and R. Kuhn, "Role-based access control," in *Proceedings of 15th National Computer Security Conference*, 1992.

[18] R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Yoman, "Role-Based Access Control Models," *IEEE Computer*, vol. 29, no. 2, pp. 38–47, February 1996.