Proactive Insider Threat Detection through Graph Learning and Psychological Context

Oliver Brdiczka, Juan Liu, Bob Price, Jianqiang Shen, Akshay Patil, Richard Chow, Eugene Bart, Nicolas Ducheneaut
Palo Alto Research Center (PARC)
3333 Coyote Hill Road
Palo Alto, 94304, CA, USA
brdiczka@parc.com

***Abstract*— The annual incidence of insider attacks continues to grow, and there are indications this trend will continue. While there are a number of existing tools that can accurately identify known attacks, these are reactive (as opposed to proactive) in their enforcement, and may be eluded by previously unseen, adversarial behaviors. This paper proposes an approach that combines Structural Anomaly Detection (SA) from social and information networks and Psychological Profiling (PP) of individuals. SA uses technologies including graph analysis, dynamic tracking, and machine learning to detect structural anomalies in large-scale information network data, while PP constructs dynamic psychological profiles from behavioral patterns. Threats are finally identified through a fusion and ranking of outcomes from SA and PP.**

The proposed approach is illustrated by applying it to a large data set from a massively multi-player online game, World of Warcraft (WoW). The data set contains behavior traces from over 350,000 characters observed over a period of 6 months. SA is used to predict if and when characters quit their guild (a player association with similarities to a club or workgroup in non-gaming contexts), possibly causing damage to these social groups. PP serves to estimate the five-factor personality model for all characters. Both threads show good results on the gaming data set and thus validate the proposed approach.

Keywords: Insider Threat Detection; Structural Anomaly Detection; Psychological Profiling; Graph Learning; Psychological Context Modeling; World of Warcraft;

I. Introduction

There is a longstanding problem of threats coming from inside government and large organizations where respected employees become malicious, giving away confidential information or even killing co-workers. These threats happen seemingly without advance notice and cause severe consequences. However, in retrospect, there is often a pattern or trail *before the fact* that could be traced and uncovered.

There are many novel technologies for detecting malicious insider behavior. Such behaviors are relatively rare in the broader user population and so techniques for anomaly detection can be applied. For example, [1] uses machine learning to recognize malicious intent in information gathering commands, [2] detects anomalies in document accesses and queries with respect to a Hidden Markov Model of text content, and [3] models user processes and flags deviations from the model. There are also many commercial tools (e.g. see, for example, [4], [5], [6]) for detecting malicious insider behavior through monitoring network activity and the use of enterprise applications.

Despite these tools, the incidence of insider attacks continues to rise in the government and commercial sectors. As an example, a recent survey found that 28% of respondents would take sensitive enterprise data to negotiate a new position in the event their employer terminated their current position [7]. Indeed, insider attacks have been the most frequent (CSI 2007, [8]) or second most frequent (CSI 2008, [9]) source of security incidents in recent years in the United States.

While these tools can accurately identify known attacks, they are necessarily reactive (as opposed to proactive) in their enforcement, and may be eluded by previously unseen, adversarial behaviors. In this paper, instead of investigation after the fact, we seek the capability to proactively identify malicious intent before the intent is carried out. Current analysis practice faces several challenges: (a) enormous amounts of data that need to be analyzed in a timely manner, (b) too many false positives in purely structural anomaly detection, and (c) the current lack of *automatic* semantic interpretation of the data at hand. We believe that it is finally possible to address these challenges in the case of adversarial insiders through the combination of two research areas: graph learning from large enterprise behavior data and psychological modeling of adversarial insiders. The first area, *Structural Anomaly Detection (SA)*, extracts information from large-scale information network data (social networks, messages, internet visits, etc.). SA defines notions of similarity between individuals, normal patterns, and anomalies, and uses technologies including graph analysis, dynamic tracking, and machine learning to detect structural anomalies in the data. The second area, *Psychological Profiling (PP)*, builds a dynamic psychological model from behavioral patterns. PP constructs psychological profiles, detects psychological anomalies and hence provides semantics for information network data. We envision a synergic interplay between the two threads. PP provides focal points for SA by filtering out a large portion that is considered irrelevant based on psychological semantics, and hence improves scalability. At the same time, PP reduces SA's false alarm rate, making the threat detection more usable and actionable.

DOI 10.1109/SPW.2012.29

This paper is structured as follows. Section II. will review relevant related work on insider threat detection, psychological modeling and graph mining. Section III. will provide details on our proposed approach including structural anomaly detection, psychological profiling, and the resulting threat fusion and ranking. Section IV. will report on some preliminary results applying the proposed approach to a data set of over 350,000 World of Warcraft (WoW) characters observed over a period of 6 month to identify malicious behavior patterns and predict player personality.

II. Related Work

Our work is related to hybrid social network analysis, insider threat detection, and psychological modeling. These are each large areas of work, hence we can highlight representative work only.

Social network analysis (SNA) is now a well-established research tool [10] with a long track record in identifying key individuals in organizations based on their communication patterns (e.g., [11]). It has been fruitfully used by the defense and intelligence community to study covert networks (e.g., [12]), in an attempt to target the most important enemies and disrupt their organization [13]. However, SNA currently suffers from three major limitations: 1) it has been used to study external threats almost exclusively, ignoring the fact that adversarial actions often come from one's own network; 2) it often ignores the psychological attributes of the actors in the network and focuses on their structural position exclusively [14]; 3) it has been restricted to static, homogeneous data sets while networks are in practice dynamic, always evolving and made of very diverse links [16].

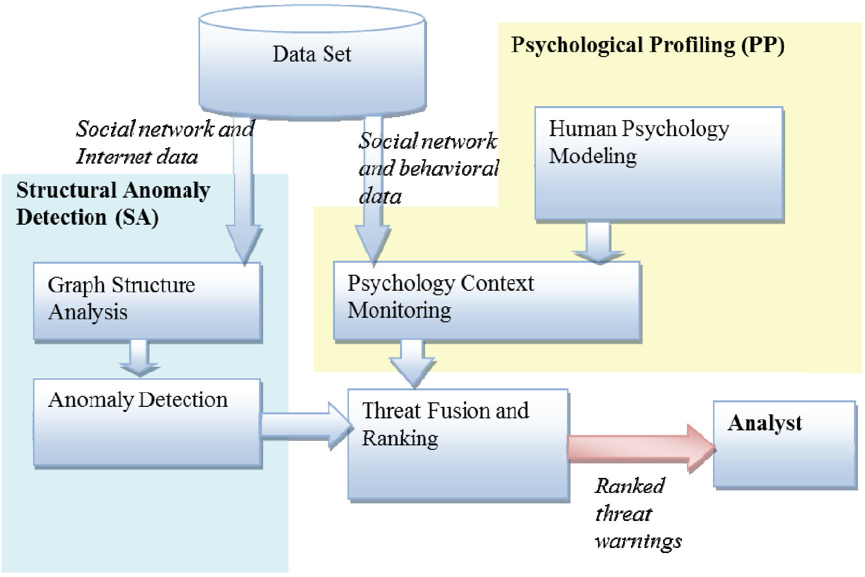
Data Set
Social network and Internet data
Social network and behavioral data
Psychological Profiling (PP)
Human Psychology Modeling
Structural Anomaly Detection (SA)
Graph Structure Analysis
Psychology Context Monitoring
Anomaly Detection
Threat Fusion and Ranking
Analyst
Ranked threat warnings

Figure 1. Architecture for detecting insider threats using Structural Anomaly Detection (SA) and Psychological Profiling (PP)

Existing techniques for detecting adversarial insiders generally focus on detecting such insiders *in the act*, as opposed to the proactive cues of the adversarial insider that are the detection goal of our framework. Malicious insider action can potentially be detected as anomalous activity on the network. For example, [1] uses machine learning to recognize malicious intent in information gathering commands, [2] detects anomalies in document accesses and queries with respect to a Hidden Markov Model of text content, and [3] models user processes and flags deviations from the model. Some augment this basic approach by introducing decoys onto the network to entrap adversarial insiders [16], [17]. A survey of much of this work is here, [18]. In addition, various models of adversarial insiders have been developed. These models include physical behaviors that are indicators of adversarial intent (e.g. foreign travel, signs of wealth) [19], as well as variables related to motivation, personality, and emotion [20], [21], [22], [23]. While all these models are valuable, none incorporate all of the possible situational triggers, context variables and indicators. We believe such attributes are necessary to establish a close connection between psychology and behavior.

When building a model linking psychological variables and adversarial insiders there is substantial psychological research to draw upon. In particular, years of research to define a taxonomy of personality attributes have led to the development of a general unifying structure of personality traits termed the five-factor model [24], [25]. The five-factor model (more commonly known as "Big 5") identifies the following general factors that represent the relationships among a host of more specific personality descriptors: Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness. In addition, more recent work has established some links between personality and behavior through models that incorporate situational triggers, context variables, and indicators in a Bayesian network designed to estimate the likelihood of future behaviors [26], [27]. Beyond these static personality factors, it is important to track enterprise behaviors that indicate changes in emotional states, in order to predict adversarial intent. There are five basic types of emotions that are recognizable across cultures: happiness, sadness, anger, fear and disgust [28]. Some studies [29] also include surprise, contempt, and interest. Because motivation and planning are necessary for the conduct of insider attacks, and these are associated with emotions, it is possible that emotional responses can provide some predictors of insider attacks. For instance, one theory of insider computer sabotage [30] suggests that an insider's anger builds to a certain level over a perceived dishonoring event, which leads to the planning of an act of revenge. Indicators of emotional states in interaction with contextual variables (such as a recent organizational change) can be used to infer thwarted insider goals and the resulting potential motivation to attack. Nevertheless, while many employees may be angry after some event, most will have personality factors that reduce the potential for extremely aggressive acts. A model of predictors of insider attack that incorporates psychological variables must therefore include both static personality traits and dynamic emotional states. Such models may build upon a dynamic model of information technology (IT) sabotage developed for the Management and Education of Risks of Insider Threat (MERIT) project [30].

III. Approach: Graph Learning for Anomaly Detection using Psychological Context (GLAD-PC)

Our approach is organized in two main technical threads (Figure 1).

The first thread, Structural Anomaly Detection (SA), extracts structural information from large-scale information network data (social networks, messages, internet visits, etc.). It defines notions of similarity between individuals, normal patterns, and anomalies, and uses technologies including graph analysis and machine learning to detect data anomalies. The SA thread contains a cascade of technical components. First, Graph Structure Analysis discovers an information network's specific characteristics and exploits them for efficient data representation and massive data reduction. Secondly, Graph Embedding converts data from a graph representation to an attribute space, which is more convenient for machine learning methods. Dynamic Tracking is performed in the attribute space to monitor individual evolution over time, and Anomaly Detection finds unusual patterns in graph data. At this point, SA detects data anomalies, not necessarily threats.

The second thread, Psychological Profiling (PP), builds a dynamic psychological model that is constructed from and informed by behavioral patterns and information network structural data. PP constructs psychological profiles and detects psychological anomalies, thereby providing semantics for the information network data. PP consists of straightforward mathematical and statistical modeling and inference: it combines a psychological model and observations (behavioral and information network structural data) to make inferences about individual's mental state and likelihood of attack. The challenge is in the psychological model --- in particular, how to make such model accurate and realistic.

We envision a synergic interplay between the two threads. PP provides focal points for SA by filtering out a large portion that is considered irrelevant based on psychological semantics, and hence improves scalability. At the same time, PP significantly reduces SA's false alarm rates, making the threat detection more usable and actionable. Likewise, Structural Anomaly Detection (SA) performs rigorous analysis of information network data, and can raise the need for psychological profiling of suspicious individuals.

Threat Fusion and Ranking then combines the data anomalies from SA and the semantic profiling from PP. Threats are ranked and presented to system users (analysts).

In the following subsections, we describe each of these building blocks in more detail.

A. Structural Anomaly Detection (SA)

Information network data often follow a somewhat repetitive pattern such as daily email communication between a manager and his employees, webpage visits of each other within a friend group in a social network, similar search and website visit patterns over time for a specific user, and so on. Identifying and generalizing these common patterns can help us define normalcy, something that accounts for the majority of the data volume that we encounter in such data sets. Under the reasonable assumption that most users in a regular social network are harmless, we are obviously not interested in the normal, but rather the rare instances of "abnormal" or unlikely patterns in the data. Graph structure analysis aims at separating the normal structures in the graph from the rare and interesting ones.

To find patterns from the nodes (individuals), we propose to leverage work on graph partitioning for community detection. The basic idea for such algorithms is to partition a graph structure so that nodes within a partition have better connectivity than nodes belonging to different partitions [31]. Existing clustering approaches range from spectral methods optimizing different partitioning criteria to heuristic-based solutions geared specifically toward particular problem instances. In this problem scenario, we are looking to identify graph structures that can be used to define normal behavior as opposed to anomalous behavior in large graph networks resembling social/information networks. The notion of an outlier community or an outlier node can be highly contextual in a graph and might also interfere with the performance of the graph-partitioning algorithm [32].

People's behavior pattern in information networks, such as whom they contact and how frequently, is inherently dynamic, typically slow-varying, with abrupt changes possibly indicating abnormal events. Methods for tracking a dynamic entity in a fixed Euclidean space (e.g., geographic location space, or abstract feature space) are well studied. For instance, sequential Bayesian methods [33] have been proposed, which recursively incorporate a temporal sequence of observation data to build up a dynamic belief in the Euclidean space. Related to the graph learning problem, we are interested in tracking nodal attributes (the "state") using temporal information network data (the "observations").

Anomaly detection assumes that subject behavior can be captured as a multi-variate sequence of abstract, continuous and discrete attributes (e.g., it might describe the evolution of attributes such as the tendency to communicate outside of the immediate group, the changing size of the peer group, the typicality of topics covered by communications, etc.). We advocate a learning-based approach, which uses behavior sequences with known anomalies to construct a probabilistic model, and then apply the model to observations to decide the likelihood of a new anomaly event. A simple baseline method, such as one-class support vector machine (SVM) or nearest-neighbor based methods, can provide a way to demonstrate early on the ability to make anomaly judgments and provide a benchmark for future performance comparisons.

B. Psychological Profiling (PP)

Psychological Profiling (PP) provides semantic meaning for threat detection. It reduces data volume, reduces false alarm rate, and detects suspicious intent/activity based on domain knowledge of human psychology. Furthermore, PP mitigates the risk that a threat could look normal from the information network data alone and would then be overlooked by the SA thread.

A psychological model gives us a way to focus our investigations on individuals who have the motivation and capability to carry out an attack. Realization of a threat requires planning and preparation, and depends on the emotional state and personality of the perpetrator. The emotional state is captured through psychological variables such as anger and depends on personal variables that represent both cognition and personality. These personal variables help to determine how external events trigger changes in emotional state, and how changes in emotional state affect the likelihood that an insider will begin or continue a threatening activity. Each of these internal model variables can be potentially associated with observable indicators. We will develop a dynamic psychological model that describes temporal patterns of activities leading up to an attack, based on relevant personality, emotional, and situational variables.

Recent research [34], [35] has uncovered relationships between enterprise work patterns and an individual's motivations, personality and emotional state. [34] and [36] collected standard personality assessments and psychometrics from enterprise workers (through a Big 5 questionnaire, see [25], and the NASA task load index, see [37]), whose work activities were logged over an extended period of time. This work identified correlations between psychometric measures (like stress/workload) and extracted temporal work patterns (like email communication patterns, document interaction patterns). Further correlations were found between personality metrics (like extraversion) and a subset of extracted temporal work patterns. In order to extract these patterns, we have developed methods and tools for analyzing streams of actions and content. The analysis comprises temporal pattern analysis of online and PC usage behavior (e.g., using T-Pattern analysis), sentiment analysis of a user's communications, and social network features and analysis.

Using standard statistical tools and techniques, we will first try to isolate correlations between one or many personality traits and other psychological variables and user behavior patterns and network data aggregated over several months and modeled as previously described. We will start with correlations between personality and high-level network variables that have been proven to be normally distributed [38], thereby making sure that any variance can be tied back to the influence of personality and psychological variables alone (the measures are density, closeness, and betweenness centrality, see [10]). We will then move on to more exploratory variables (e.g. k-cores, cliques), with the understanding that results could be less robust.

In addition to tracking static behavior patterns and social network structures, we also aim to detect sudden changes in behavior and social network configuration. Sudden changes may be indicative of an imminent insider attack. For example, as noted earlier, many spies underwent a major personal loss before engaging in espionage [39]. Therefore, a sudden change in emotions or other psychological variables is likely to be associated with the start of an attack. Changes in emotions or other psychological variables are often manifested in the daily activities (and changes therein) of the adversary.

C. Threat Fusion and Ranking

As explained in the previous subsections, SA detects abnormal patterns in information network data and PP constructs psychological profiles from behavioral data. We propose a Bayesian method for Threat Fusion and Ranking, to coherently combine data about psychological anomalies and profiles, and provide threat alarms. With semantics provided by PP, we believe we can reduce the false alarm rate significantly. We propose a Bayesian fusion method, which assumes a generative model for behavioral and graph anomalies, i.e., if an individual has malevolent intent, how the intent may show up in behavior and information network data. Given an accurate generative model and the anomaly observations (from SA and PP), this method infers the underlying threat. An initial generative model can be generated from limited historical data like the Fort Hood event, and a set of example scenarios covering threats and innocuous behaviors.

Statistical inference method will rank threats based on their probability and uncertainty. In addition, in order to provide actionable information, ranking should also take into account other factors, such as the severity and urgency of threats. We should also consider the cost of errors, i.e., missed detections and false alarms. Missed detection may be extremely expensive, e.g., human lives lost, while false alarms can inflict unfair discrimination on people. When presenting potential threat to users (i.e., military or safety personnel), ranking should be designed carefully to incorporate cost metrics, e.g., average, and/or worst case costs. We propose to combine statistical inference methods with decision theory to achieve optimal or near-optimal ranking solutions.

IV. Application: Detecting Malicious Insiders in World Of Warcraft (WoW)

This section describes the application of the proposed approach on a large data set coming from online gaming. We would like to stress that the results reported here are work in progress and are intended to illustrate the proposed approach. The motivation to use an online gaming data set relies in the fact that there is only a very limited amount of data available showing behaviors of malicious insiders. In contrast, enormous amounts of online gaming behavior traces are publicly accessible, and due to the anonymity of game characters, the data does not have similar privacy restrictions as other data (e.g., from an enterprise context). World of Warcraft (WoW) is the most popular US-based massively multi-player online game with 10.3 million subscribers as of November 2011 [40], which is the reason why we chose WoW as basis for the analyses in this paper. We use data from a previous study [41] that exhibits different types of malicious behaviors. A web-based crawler was deployed to log in-game activities based on the API specified by Blizzard Entertainment, the producer of WoW. The crawler periodically issues ”/who” requests every 5 to 15 minutes, depending on server load, to get a list of characters currently being played on a given server. Over six months of data are logged, from November 2010 to May 2011. The data is sometimes referred to as the *WoW census*. Data was logged from three different servers: Eitrigg, Cenarion Circle, and Bleeding Hollow. Overall we observed more than 470,000 unique characters forming over 15,000 guilds. A guild is a group of game characters (ranging from two to hundreds of members) who voluntarily chose to share resources, plan strategy and join forces in large-scale group events against computer-controlled opponents (“bosses” in dungeons) or players from the opposite faction (indeed, each server is divided into two factions in a permanent state of war: the Alliance and the Horde; players must choose one side when they create their character). In addition to the WoW census, 1,040 game players were recruited to participate in a study assessing their personality. Participants were asked to list up to 6 WoW characters they were actively playing. This resulted in a total of 3,050 active characters. Participants completed a web-based survey that gathered their demographic and personality information. A 20-item survey measuring the Big 5 factors was drawn from the International Personality Item Pool. Participants rated themselves on the items using a scale ranging from 1 (Very Inaccurate) to 5 (Very Accurate).

In order to illustrate our approach, we investigate and report on two distinct parts: detecting destructive group dynamics (using methods from Structural Anomaly Detection (SA)), and inferring online gamer personality (using methods from Psychological Profiling (PP)). Both approaches will be described in the following.

A. Detecting Destructive Group Dynamics

Guilds are a vital part of the social life of WoW, providing a basic unit for social networking that makes the gaming experience appealing and enables players to take on difficult tasks as a team. The effectiveness of these groups can be undermined when group members depart, taking with them experience, resources and possibly other group members. The ability to predict imminent departure of group members and the probable impact of their departure is highly desirable, as it offers insights on factors that affect online group cohesion and effectiveness.

Social interaction may be an important influencing factor in guild-quitting events. First, we define a friendship network among guild members, where nodes are characters, and edges indicate co-occurrence within gaming zones — if two characters were observed in the same game location (zone in WoW), an edge is added between the corresponding nodes. The underlying assumption is that if characters co-occur in a gaming zone, it is highly likely that the characters are collaborating on a gaming activity. We further add a membership network to indicate the affiliation between characters and guilds. Nodes fall into two categories: (1) guild nodes, and (2) character nodes. If a character is observed appearing in a guild, an affiliation edge is added. The overall network is the super-imposition of the friendship and the affiliation networks. It is an undirected multi-graph i.e. it allows for multiple edges between any two nodes in the network.

Table 1 lists some statistics in the social network from the WoW census. The first thing that one might notice is that we lose a significant number of characters when constructing the network from the raw data: for instance, the server Eitrigg contains 176,166 characters, but only 51,244 characters show up in the social network. On further study, the missing ones are typically lower level character in the game (it is important to note that collaboration in the game begins at medium to higher levels) and with a very small footprint of existence during the six months of our crawl. These characters are essentially low-level “alts”, characters created by the players to explore other “classes” from the game, which have been abandoned after a few levels (e.g. a player might want to try out a warrior after having played a wizard for a few months, to see if this class fits his or her play style better, but then choose not to develop this “alt” any further). These characters can also be “mules”, low-level characters created purely to stockpile items. Both can be safely ignored as our interest is in the analysis of social group dynamics. Also, we note that guild-quitting events are fairly common — around 26% of characters quit from a guild at least once in our observation period on Eitrigg. Similar guild quitting statistics are observed on Cenarion Circle and Bleeding Hollow.

Table 1. Overall Network Statistics

Statistic	Eitrigg	Cenarion Circle	Bleeding Hollow
# of characters	51,224	72,108	47,499
# of guilds	2906	3425	2911
# of edges	2,447,577	3,713,390	2,827,789
% Characters changing guild	26.53	32.28	20.69

An important question to ask is whether there exists a snowballing effect in WoW with respect to quitting events in a guild, or, on the contrary, quitting events are independent. The existence of a snowballing effect would mean that a quitting character would be able to cause damage to or even destabilize a guild. In order to investigate this, we test whether the quitting events in a guild follow a Poisson process, a well-known probabilistic model for discrete events arrivals. Table 2 summarizes the result of the goodness-to-fit test. A large fraction of the quitting events do *NOT* follow a Poisson process. This indicates that quitting events are not independent, but rather correlated.

Table 2. Guild quitting event analysis

Server	Number of Guilds Analysed	Following Poisson Process	NOT Following Poisson process
Eitrigg	181 (6.23%)	23 (12.71%)	*158 (87.29%)*
Cenarion Circle	102 (3.50%)	28 (27.45%)	*74 (72.55%)*
Bleeding Hollow	309 (9.02%)	28 (9.06%)	*281 (90.94%)*

Guild quitting events being correlated, we investigate whether it is possible to predict *if and when* a character will

Table 3. Results for quitting event prediction

Server	Overall Accuracy	Non-Quitting Events			Quitting Events		
		Precision	Recall	F-Measure	Precision	Recall	F-Measure
Eitrigg	82.7432	0.878	0.926	0.901	0.389	0.268	0.317
Cenarion Circle	89.0973	0.917	0.967	0.941	0.342	0.164	0.222
Bleeding Hollow	79.8396	0.855	0.91	0.811	0.396	0.276	0.325

quit his or her guild. The aim here is to build a useful detector for identifying characters who might potentially quit and abscond with guild resources or otherwise damage the effectiveness of the guild. The prediction problem is formulated as the following: given the game trace up to current time, predict whether a character will quit from the guild within a specified future interval. Game and social network events within a guild are grouped by day, called a "day record". Based on this notion, we define a prediction window in terms of day records (7 in our experiment). If this window of future day records includes observations of the character in another guild, we set the class label *target_guild_has_changed* to true to indicate a quitting event has occurred. This class label is our prediction target.

Prediction of quitting event is inherently dynamic. Game activities unfold as time advances, and prediction should be updated accordingly. To accommodate the dynamic nature, we keep a running window of features. A personal history window of 14-day records is used to generate features summarizing the character's playing history including game engagement, playing time and achievements within the history window. In addition, social features regarding the guild, especially social relationships with guild members, are used. These include clustering coefficient (measuring structure balance), playing time within the guild and collaboration time (measuring engagement within the guild).

In order to reduce the risk of overfitting of our anomaly detection/prediction, we developed distinct training and test sets where history and prediction windows in the training set are generated from a different set of characters than the history and prediction windows in the test set. Additionally, as quitting events are rare by comparison, we use a random sampling method to balance the training set so that it contains approximately equal proportions of non-quitting and quitting events.

Table 3 reports the classification performance for a random forest with 10 trees and unlimited depth and feature counts. Guild quitting prediction classifiers are built separately for the 3 WoW servers. Classification performance does not vary much on servers, indicating that our approach is generally applicable in the WoW space.

We conclude that it is possible to predict quitting events with modest precision and recall and that past loyalty, guild stability, and social engagement are key predictive factors.

B. Inferring Online Gamer Personality

A person's personality affects his/her activities and produces many cues. The goal of the personality prediction in WoW is to estimate the personality of the player that is behind a character in the game from observable in-game behavior or characteristics (e.g., achievements, or social network features). In this paper, personality is represented by factors from the Big 5 model consisting of extraversion, agreeableness, conscientiousness, neuroticism, and openness. The previously mentioned data set of 3,050 active characters (from 1,040 game players) including their WoW census data and Big 5 profiles from an online questionnaire is the basis for training and testing our approach.

We analyze a game character's personality based on 3 kinds of sources – behavioral, text analysis and social networking information. First, we note that different people choose different strategies to achieve their objectives in the game, which lead to different behaviors. For example, shy and quiet players might prefer solo activities such as cooking and fishing (these individual "professions" produce food that can be used to provide "buffs" to characters, that is, temporary improvements of their abilities). Similarly, players who score low on extraversion could have a preference for some player-vs.-player activities (duels, arenas, battlegrounds). We therefore analyze a character's in-game activity based on 68 behavioral features including achievements (given for reaching important milestones in the game), different types of deaths (e.g., falling versus combat), "respecs" (changes in the mix of a character's abilities, for instance to emphasize attack or defense), character skills, use of emotes, equipment, pets (cute companions with no functional value in the game), balance of game activities (e.g. large-scale dungeons vs. individual quests), and finally damage and healing done to other characters.

Furthermore, we note that a player usually picks a character name appealing to him/her after some thinking. Guild names are also chosen carefully to reflect the intended "social tone" of the group (for instance, "Merciless Killers" conveys a different impression than "The Merry Bards"). We therefore analyze character names and the name of a character's guild to get additional information on the player's personality. To do so, we utilize sentiment analysis, keyword lists and n-grams to generate features. We apply sentiment analysis to character names and guild names using two sentiment polarity dictionaries for this purpose. We scan the name and count how many positive/negative/neutral/both words appear. We further created a game keyword list and check if the name contains those keywords. Those keywords include race names (such as elf, gnome), role names (such as priest, warrior), actions (such as kill, wave), reasons for failure (such as drown, fatigue), in-game activities (such as arena, dungeon) and other frequent words. We currently collect 80 keywords. In the textual analysis domain, n-gram analysis is a popular technology that uses sliding window character sequences in order to aid classification. To capture other hidden patterns in character and guild names, we also construct n-grams from names. An n-gram is a subsequence of n letters from a given sequence. For example, if the character name consists of 4 letters – ABCD, then we will have bigram AB, BC, CD. We limit n to 4, i.e., we only consider bigrams, trigrams and 4-grams. Larger n gives too much computation complexity and does not improve accuracy much. In many cases, the character name is related to the game player's other choices in the virtual world, such as race and gender. Thus we include the character's region, virtual gender, race, role and faction as additional features. We train regression trees to profile personality from the above text information.

We also hypothesize that a person with some specific personality traits (for instance, extraversion or agreeableness) is more socially active. We analyze the friendship and membership network described in the previous section in order to generate predictive features in that domain. We extract graph characteristics for each node (person) in the network such as degree centrality, betweenness centrality, and closeness centrality. We then calculate their maximum values, minimum values and histograms as our features and input them into regression trees for profiling personality.

Each of the predictors above provides a partial and complementary view of personality. As a last step, we fuse these individual predictions together to get a more accurate personality profile through linear regression in order to reach higher accuracy (Figure 2).

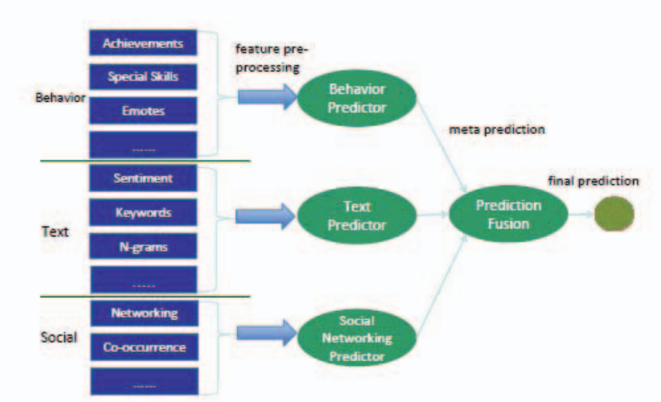

Figure 2. Several predictions are combined to get the final prediction of personality.

As personality variables (and their predictions) are real numbers, we evaluate the approach using the Pearson correlation between the real and predicted values. The obtained results (Figure 3) are based on 10-fold cross-validation. Text analysis and features seem to be the most predictive indicator of a person's personality (with correlation values between 0.25 and 0.4). By fusing predictions from multiple views together, we obtain the best results.

C. *Next Steps*

As mentioned in the introduction to this section, the results based on the WoW data reported here are work in progress, and the construction of the (complete) analysis pipeline for this data (as sketched in Figure 1) is ongoing. We obtained promising results on Structural Anomaly Detection (detecting destructive group dynamics in guilds) and Psychological Profiling (inferring gamer personality from fused features). Each of those two components will output a ranked list of individuals/characters. The ranking will be based on the maliciousness of their actions (guild quitting and damage) or deviating (predicted) personality variables. The next step is to combine each of these rankings in a (yet to be built) threat fusion and ranking component. We envision the component to use a generative model that is optimized against suitable metrics (e.g., human perceived severity of a threat) coming from the end-users of the system (analysts).

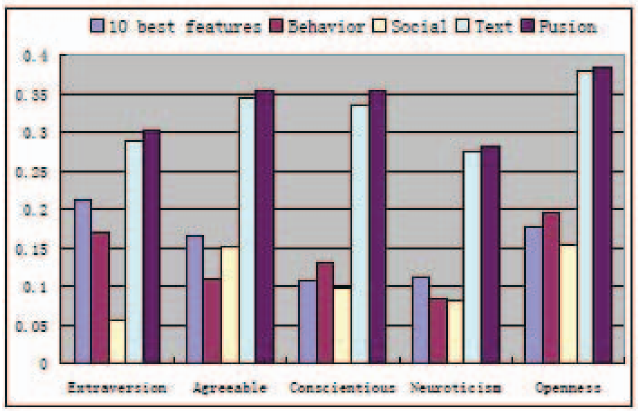

Figure 3. Correlation of personality predictions with real values for behavior, social network, text and fused features.

V. CONLUSION

We have presented an approach for proactive detection of insider threats by combining structural anomaly detection from social and information networks and psychological profiling of individuals. The approach has been tested on a large data set from the massively multi-player online game World of Warcraft including over 350,000 game characters observed over a period of 6 months. In contrast to many real-world data sets and collections, the game data contains obvious malicious behaviors that are identifiable and are not constraint by any legal or privacy regulations (game data is publicly accessible). First obtained results indicate that is it possible a) to detect anomalous behaviors (guild quitting) through structural analysis of social networks in the game, and b) to predict a player's personality from in-game behavior and other in-game features. While the game data sets help validate our approach both in terms of scalability and feasibility, we believe that the proposed approach can similarly be applied to large-scale enterprise data sets including communication data like email or IM, file accesses, and login data. Our next steps will concern further work on threat fusion and ranking (including the presentation to the end-users/analysts) and the application of the approach to a large-scale enterprise behavior data set that is currently being collected.

ACKNOWLEDGMENT

The authors gratefully acknowledge support for this work from DARPA through the ADAMS (Anomaly Detection At Multiple Scales) program funded project GLAD-PC (Graph

Learning for Anomaly Detection using Psychological Context). Any opinions, findings, and conclusions or recommendations in this material are those of the authors and do not necessarily reflect the views of the government funding agencies.

The authors would also like to thank GLAD-PC team members Elise Weaver and Paul Sticha of HumRRO for help in understanding the psychology of adversarial insiders.

REFERENCES

[1] M. Salem and S. Stolfo. Masquerade attack detection using a search-behavior modeling approach. Columbia University Computer Science Department, Technical Report # cucs-027-09, 2009.

[2] P. Thompson. Weak models for insider threat detection. Proceedings of the SPIE Vol. 5403, Sensors and Command, Control, Communications and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense III, 2004.

[3] P. Bradford and N. Hu. A layered approach to insider threat detection and proactive forensics. ACSAC 2005.

[4] Raytheon Oakley Systems SureView. http://www.raytheon.com/capabilities/products/cybersecurity/insiderthreat/products/surview/ (Retrieved Feb 14, 2012)

[5] Lanxoma. Intelligent Desktop Surveillance. http://www.lanxoma.com/ (Retrieved Feb 14, 2012)

[6] PacketMotion. http://www.packetmotion.com/ (Retrieved Feb 14, 2012)

[7] Cyber-Ark Press Release, November 23,2009. http://www.cyber-ark.com/news-events/pr_20091123.asp (Retrieved Feb 14, 2012)

[8] Computer Security Institute (CSI) Computer Crime and Security Survey, 2007. http://i.cmpnet.com/v2.gocsi.com/pdf/CSISurvey2007.pdf (Retrieved Feb 14, 2012)

[9] Computer Security Institute (CSI) Computer Crime and Security Survey, 2008. http://www.docstoc.com/docs/9484795/CSI-Computer-Crime-and-Security-Survey-2008 (Retrieved Feb 14, 2012)

[10] Wasserman, S. and K. Faust (1994). Social network analysis: methods and applications. Cambridge, UK, Cambridge University Press.

[11] Burt, R. (1995). Structural holes: the social structure of competition, Harvard University Press.

[12] Sparrow, M. (1991). ”The application of network analysis to criminal intelligence: an assessment of the prospects.” Social Networks 13: 251-274.

[13] Carley, K. (2003). Destabilizing terrorist networks. 8th International Command and Control Research and Technology Symposium, National Defense War College, Washington DC.

[14] Burt, J. E. and J. T. Mahoney (1998). “Personality correlates of structural holes.” Social Networks 20: 63-87.

[15] Carley, K. (2003). Dynamic network analysis. NRC workshop on social network modeling and analysis.

[16] L. Spitzner. Honeypots: Catching the insider threat. ACSAC 2003.

[17] B. Bowen, S. Hershkop, A. Keromytis and S. Stolfo. Baiting inside attackers using decoy documents. SecureComm 2009.

[18] M. Salem, S. Hershkop and S. Stolfo. A survey of insider attack detection research. Insider Attack and Cyber Security: Beyond the Hacker, Springer, 2008.

[19] M. Marbury, P. Chase, B. Cheikes, D. Brackney, S. Matzner, T. Hetherington. B. Wood, C. Sibley, J. Marin, T. Longstaff, L. Spitzner, J. Haile, J. Copeland and S. Lewandowski. Analysis and Detection of Malicious Insiders. Technical Paper, Case #05-0207.

[20] K. Herbig. Changes in espionage by Americans: 1947-2007. Department of Defense Technical Report 08-05, March 2008.

[21] K. Herbig and M. Wiskoff. Espionage against the United States by American citizens 1947-2001. PERSEREC Technical Report 02-5, July 2002.

[22] S. Band, D. Capppelli, L. Fischer, A. Mooore, E. Shaw and R. Trzeciak. Comparing Insider IT Sabotage and Espionage: A Model-Based Analysis. Technical Report CMU/SEI-2006-TR-026.

[23] M. Keeney, E. Kowalski, D. Cappelli, A. Moore, T. Shimeall, S. Rogers. Insider Threat Study: Comuter System Sabotage in Critical Infrastructure Sectors. U. S. Secret Service and CERT Coordination Center/SEI.

[24] Digman, J.M. (1990). Personality structure: An emergence of the five-factor model. Annual Review of Psychology, 41, 417-440.

[25] Costa, P.T.,Jr. and McCrae, R.R. (1992). Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) manual. Odessa, FL: Psychological Assessment Resources.

[26] Sticha, P.J., Buede D.M., and Rees, R.L. 2005. APOLLO: An analytical tool for predicting a subject’s decision making. 2005 International Conference on Intelligence Analysis Proceedings. Bedford, MA: the MITRE Corporation.

[27] Sticha, P.J., Buede D.M., and Rees, R.L. 2006. Bayesian model of the effect of personality in predicting decisionmaker behavior. Proceedings of the 22nd conference on uncertainty in artificial intelligence.

[28] Ekman, P. (1973). Cross-cultural studies of facial expression. In P. Ekman, (Ed.), Darwin and facial expression: A century of research in review. NY: Academic Press.

[29] Ekman, P.&Friesen,W.V. (1971). Constants across culture in the face and emotion. Journal of Personality and Social Psychology, 17, 124-129.

[30] Cappelli, D.M., Moore, A.P., & Shaw, E.D. (2006). A risk mitigation model: Lessons learned from actual insider sabotage. Presentation to the Computer Security Institute, Orlando, FL, November 7, 2006. Downloaded from http://www.cert.org/archive/pdf/CSInotes.pdf (Retrieved Feb 14, 2012)

[31] Leskovec, J., Lang, K., and Mahoney, M. “Empirical Comparison of Algorithms for Network Community Detection”, ACM WWW International conference on World Wide Web (WWW), 2010.

[32] Gao, J., Liang, F., Fan, W., Wang, C., Sun, Y., and Han, J. (2010) “Community Outliers and their Efficient Detection in Information Networks”, ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD).

[33] Arulampalam, M. S. , Maskell, S., Gordon, N., and Clapp, T. (2002) “A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking”, by IEEE Trans. on Signal Processing, Vol 50, No. 2, Feb 2002.

[34] Brdiczka, O., Su, N., and Begole, B. (2009) “Using temporal patterns (t-patterns) to derive stress factors of routine tasks”, CHI Extended Abstracts.

[35] Olguin, D., Waber, B.N., Kim, T., Mohan, A., Ara, K., and Pentland, A. (2009) “Sensible Organizations: Technology and Methodology for Automatically Measuring Organizational Behavior”, IEEE Transactions on Systems, Man, and CyberneticsPart B: Cybernetics, Vol. 39, No. 1, February 2009.

[36] Su, N., Brdiczka, O., and Begole, B. (2012) “The routineness of routines”, to appear in HCI Journal.

[37] Hart, S.G. , and Staveland, L.E. (1988) “Development of nasa-tlx (task load index): Results of empirical and theoretical research”, Human Mental Workload 1, 139183.

[38] McCulloh, I. and K. Carley (2008). “Social network change detection - Technical Report CMU-ISR-08- 116”, Carnegie Mellon University, School of Computer Science, Institute for Software Research.

[39] Herbig, K., and Wiskoff, M. (2002) Espionage against the United States by American citizens 1947-2001. PERSEREC Technical Report 02-5, July 2002.

[40] F. Cifaldi. World of warcraft loses another 800k in three months. gamasutra.com, November 2011.

[41] Ducheneaut, N.; Yee, N.; Nickell, E.; and Moore, R. J. (2007) The life and death of online gaming communities: A look at guilds in world of warcraft. Prof. of CHI.