Safeguarding Academic Accounts and Resources with the University Credential Abuse Auditing System

Jing Zhang[1], Robin Berthier[2], Will Rhee[3], Michael Bailey[1], Partha Pal[4], Farnam Jahanian[1], and William H. Sanders[2]

[1]Dept. of Electrical Engineering and Computer Science, University of Michigan
{*jingzj, mibailey, farnam*}*@umich.edu*
[2]Information Trust Institute and Dept. of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign
{*rgb,whs*}*@illinois.edu*
[3]Office of Information and Infrastructure Assurance, University of Michigan
willrhee@umich.edu
[4]BBN Technologies, Cambridge, MA
ppal@bbn.com

***Abstract*—Whether it happens through malware or through phishing, loss of one's online identity is a real and present danger. While many attackers seek credentials to realize financial gain, an analysis of the compromised accounts at our own institutions reveals that perpetrators often steal university credentials to gain free and unfettered access to information. This nontraditional motivation for credential theft puts a special burden on the academic institutions that provide these accounts. In this paper, we describe the design, implementation, and evaluation of a system for safeguarding academic accounts and resources called the University Credential Abuse Auditing System (UCAAS). We evaluate UCAAS at two major research universities with tens of thousands of user accounts and millions of login events during a two-week period. We show the UCAAS to be useful in reducing this burden, having helped the university security teams identify a total of 125 compromised accounts with zero false positives during the trail.**

***Keywords*-compromised account, university, authentication, Virtual Private Network (VPN)**

I. Introduction

Data theft on the Internet is a booming business [1]. Analysis of phishing [2] and malware [3] incidents shows that attackers are very interested in financial data, with banking sites routinely topping the list of targets for data theft. As a result, institutions and researchers have dedicated considerable attention to addressing the issue of stolen and compromised financial credentials. While it is not surprising that university environments have their own share of credential theft, an analysis performed at our own academic institutions reveals that credentials stolen in university settings are typically used not to acquire financial data, but rather *to gain free and unrestricted access to information.* Compromised accounts actively utilize Virtual Private Network (VPN) and library publication resources. The VPN enables attackers to bypass censorship mechanisms deployed in their countries. Recently, the exploitation of scholarly databases has also become a lucrative business, as attackers download a large number of articles and then resell them on underground markets [4].

To address the problem, we present in this article the design, implementation, and evaluation of a VPN abuse detection system that focuses on supplementing existing security measures to rapidly identify account compromises. Our system analyzes authentication logs on a daily basis and reports accounts for which suspicious activity is detected. The detection technology is based on a machine-learning approach that automatically generates a set of features before classifying user activity. Our work makes three important contributions. First, we report on the motivation of attackers who compromise academic accounts based on several years of incident analysis at two large universities. Second, we present the design of an authentication log analysis solution that can process the daily activity of thousands of accounts with high accuracy and a low false-positive rate. Third, we evaluate this system on several weeks of logs at each university in close collaboration with the institutions' security teams. This large-scale experiment has led to interesting insights about the specific challenges of analyzing malicious activity from campus data.

II. Background

University accounts have become an attractive target for attackers seeking access to online campus resources. In fact, during 2010 and the first six months of 2011, the security team at the University of Michigan (UofM) recorded 613 incident tickets related to unauthorized use of university accounts. The incidents represented a significant fraction of the workload of security officers. The issue was not limited to UofM; the security team at the University of Illinois at Urbana-Champaign (UIUC) recorded tickets for 26 compromised accounts in the first half of 2011. This burden led security teams at both universities to engage researchers in an effort to better understand the threat and help determine how to mitigate it.

A. University Account Compromises

A study of the account compromises revealed a variety of attack motivations. Traditional malicious activity included spamming via university email accounts or installation of malware on public machines. Adversaries also attempted to access confidential user information or even access confidential databases (e.g., hospital records, human resource databases) via authorized accounts. Surprisingly, some compromise incidents involved users who simply wanted to access seemingly mundane resources in an effort to have free and unfettered access to information. For example, some accounts were used to gain access to scholarly publications available for free from university IP addresses, or to circumvent Internet filtering in the attacker's country of origin.

To understand the activity in more detail, the security team at UofM collected 24 hours of Netflow data for the VPN sessions of 8 compromised accounts. The accounts were verified as compromised by the UofM security team, and the 8 selected accounts were confirmed by the security personnel to have no legitimate activity. All 8 accounts were accessed from China. Our first observation was that the illegitimate users visited ten websites (e.g., Facebook, YouTube) blocked by the network censorship deployed in China. We confirmed these filters via Chinese end hosts. Those activities accounted for 8.2% of the total HTTP Netflow we observed. Second, the library website was repeatedly visited during these sessions. A third interesting observation was that hundreds of Netflow records (5.6% of the total HTTP records) showed the adversaries logging in to 7 accounts at other universities, which we conjecture was a strategy to evade detection.

Figure 1. Chinese online market where account pirates sell stolen credentials for publication download services. Here, year's access to journals at multiple scholarly databases is sold for 500RMB (less than $100).

Interestingly, we found that such illegitimate VPN and library access was the most common kind of compromised account activity (18.9% of the total incidents at UofM). Of course, this type of access is not merely used by the attackers themselves, but often resold. In June 2011, it was revealed that stolen accounts from college students and professors were sold on Taobao, one of the largest Chinese online marketplaces, giving access to the SciFinder database of scientific articles [4]. Stolen credentials were provided for less than ten dollars per month, complete with instructions on how to use the university VPN and access scholarly resources. Other sellers advertised that for a few pennies they would provide scientific publications on-demand within 24 hours. Fig. 1 shows a screenshot of such an advertisement from Taobao, an online market in China.

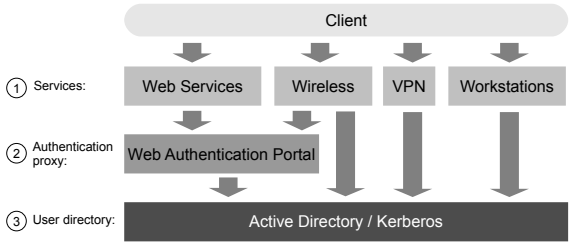


Figure 2. Authentication infrastructure.

B. University Authentication Infrastructure

Universities usually provide an abundance of resources and online services to vast and diversified user groups at both the campus and college levels. For example, users have access to webmail, VPN, course registration, online storage, payroll and employment information management, and library resources. According to data collected in November 2010, there are 41,924 students and 34,947 faculty and staff members (including university hospital personnel) at UofM, but the total number of unique accounts reaches 556,281 because alumni and former employees can continue to use their accounts after they leave the campus. Among those accounts, 206,529 had recent activity. At UIUC, the population includes 54,612 people, including 43,862 students and 10,750 faculty and staff members. Unlike UofM, UIUC locks accounts a few months after students graduate or employees leave the university.

As shown in Fig. 2, UofM and UIUC share a similar authentication infrastructure built around the Kerberos authentication protocol and different authentication portals for various services. One is the single-sign-on protocol for Web services (Cosign at UofM and Bluestem at UIUC). Both Cosign and Bluestem are browser-based Web authentication solutions that enable users to access restricted online resources or Web applications at the universities. The other important authentication channel is the VPN service. Users can gain remote access to the university network through different VPN clients. Logs from all the VPN gateways are recorded. Those two types of service are our main focus because they represent the majority of campus usage and account compromises are likely to manifest in their logs.

III. Detection System

A. Overview

1) Goals: The main design goal of the University Credential Abuse Auditing System (UCAAS) is to assist secu-

rity teams by automatically flagging compromised accounts. The first requirement is that the system must be effective in detecting compromised accounts, even if the illegitimate activity is stealthy. The effectiveness of UCAAS relies heavily on the account activity features chosen for classification. We engaged in discussions with the security teams at both universities, and we examined authentication logs from the past two and a half years in order to gain a detailed understanding of legitimate and malicious activity. As a result, we carefully selected a combination of features in which illegitimate activity would be most likely to manifest. We describe our selection in Section III-B and evaluate it in Section IV-B. Second, given the relative scarcity of compromised accounts compared to the total population, the distribution of false positives over true positives is likely skewed [5]. The cost of false positives impacts both the operational team, which has to spend time investigating flagged accounts, and legitimate users, because their accounts could be blocked. Discussion with the security teams helped us to define the requirement of a maximum average number of *two false alarms per day*.

2) Overall Design: UCAAS detects suspicious accounts based on authentication logs collected from university systems. Since VPN is used as an entry point by attackers, we first filter authentication activity to keep only logs generated by users who accessed the VPN at least once. The second step consists of extracting and analyzing the set of features for the daily activity of each account. The activity captured by the different features describes the behavioral, geographical, and topological pattern information, as well as possible deviations from historical profile data. Finally, a classifier runs on the feature vectors to determine whether an account is compromised or not. The model used by the classifier is trained and built automatically from the past n days of authentication logs. This training dataset is made of both known compromised and legitimate accounts, and is updated dynamically over time.

B. Features

1) Suspicious Behavior Features: First, we created three heuristics based on our analysis of compromised account behavior patterns. The thresholds in the heuristics are transparent, as they were derived from operator experience.

Temporal-Spatial Violation: This feature captures accounts that had activities from geographically different locations in a short period of time. The key insight here is that the activities of legitimate account owners and attackers are independent. Thus, if attackers and account owners are located in different places, they will likely generate inconsistencies in location and time of activities. Since attackers can neither predict nor control the legitimate users, this detection can hardly be evaded.

Suspicious IP Addresses: UCAAS labels an IP address as suspicious if it was used by more than one account to log in during one day. An analysis of historical compromised account authentication logs revealed that a handful of IP addresses scanned multiple compromised accounts within a short period of time. In fact, at UofM, more than half of the compromised accounts reported in 2009 and 2010 were accessed by those malicious IP addresses.

Suspicious Usage Pattern: Account activity is labeled as suspicious if it consists exclusively of a combination of VPN and library accesses. Unlike the first feature, which relied on overlapping legitimate and malicious activities, this feature aims to detect idle accounts that are used only by the compromising entity. For example, alumni and student accounts are mostly inactive during vacation periods; once those accounts have been compromised, their usage patterns no longer reflect traditional academic activities (e.g., registering for classes, connecting to the wireless network). Note that we use a threshold to characterize how exclusive the usage pattern should be before an alert is raised. This is to prevent attackers from evading the heuristic by randomly logging in to resources other than the VPN and the library.

While this first group of features focuses on defining specific signatures for illegitimate activity, the remaining three groups of features take a complementary approach by learning anomaly-based characteristics over time.

2) IP Address-based Features: We observed that the geographical and topological distributions of attackers were not uniform. As a result, UCAAS generates three IP address-based features to capture client origins that are likely linked to illegitimate activity: *geographic location*, *Autonomous System Number (ASN)*, and *Top-Level Domain (TLD)*.

3) Resource Usage-based Features: To expand on the suspicious usage pattern heuristic, the system learns the ratios of resource usage for the account. These ratios are computed for each account based on the numbers of VPN connections and university website accesses. This set of features is based on the intuition that legitimate users usually go to a variety of online services provided by their universities, while attackers only exploit a few services heavily.

4) Profile-based Features: Finally, for each account, UCAAS computes the probability that the latest activity recorded matches the historic profile of user activity collected over the past week. An example profile is presented in Table I. If we observe a new authentication attempt from the United States for this sample account, the probability that this attempt fits in the historical usage pattern is 91.2%. The idea behind these features is that illegitimate activity will not match the authentication habits of account owners.

C. Classification

The next step after computing values for the different features is to classify accounts into two classes: benign or suspicious. For this task, UCAAS uses a logistic regression classifier implemented in Weka [6]. We initially ran experiments with a variety of machine-learning algorithms, including support vector machine, naive Bayes, and K-nearest neighbors algorithms. Most of them offered good results if the parameters were correctly tuned. Overall, we found that

	Timing-related	
Feature	*Time of the Day*	*Day of the Week*
Value	00:00–04:00: 3% 04:00–08:00: 7% 08:00–12:00: 30% 12:00–16:00: 40% 16:00–20:00: 15% 20:00–00:00: 5%	Monday: 17.2% Tuesday: 12.2% Wednesday: 21.0% Thursday: 19.0% Friday: 20.0% Saturday: 7.8% Sunday: 2.8%

	Location-related		
Feature	*TLD*	*ASN*	*Country Code*
Value	.COM: 86.9% .NET: 9.1% .ORG: 4.0%	27432: 79.6% 123: 20.4%	US: 91.2% CN: 8.8%

	Resource-related
Feature	*Usage Frequency*
Value	VPN: 22.3% Wireless: 77.7%

Table I
SAMPLE USER PROFILE LEARNED OVER A WEEK OF ACTIVITY

a logistic regression classifier provided the best accuracy. Indeed, a logistic regression model [7] is inherently suitable for single dichotomous label classification. The model is $L = \sum B_i X_i$, where each feature X_i has a coefficient B_i. For each incoming feature vector, the classifier calculates the natural logarithm L of the odds that a compromise happens: $L = \ln \frac{\hat{p}}{1-\hat{p}}$, where $\hat{p}$ represents the estimated probability that the account is compromised. The higher L is, the more likely it is that the account has been compromised. The final classification uses a threshold to identify accounts as benign or possibly compromised.

IV. EVALUATION

A. Datasets and Ground Truth

We used two datasets to evaluate UCAAS: a *training set* used for feature tuning and model testing, and a *validation set*. One critical and difficult step in building the training set is that of obtaining the ground truth. We addressed this challenge through a close collaboration with the security teams, which have acquired extensive experience in dealing with compromised accounts over the past several years. The first step was to collect known incident tickets from 2009 to 2011. They represent a subset of the total compromised accounts, since many compromised accounts are never identified. Therefore, we needed further manual checking of the dataset. However, the large volume of logs in our dataset (around 6 million) made the manual validation of each authentication attempt impractical. To address this problem, we ran the heuristics discussed in Section III-B1 with conservative parameters tuned to minimize false negatives. The set of flagged accounts was sent to the security team. They examined each account and contacted the owners of suspicious ones via email and telephone to assess whether or not the suspicious activity detected was really illegitimate. Most of the time, this validation step consisted of asking if the user traveled to the foreign country identified in the data, or if the user shared credentials with other people. For alumni or previous employees at UofM who could not be reached by email or telephone, the security team provided an expert judgment. The final step was to manually examine the authentication activities of those compromised accounts, and label their feature vectors as compromised in the days during which we had high confidence that illegitimate activity had occurred. We kept refining the ground truth with the latest detection results during the course of the evaluation process. We believe that through this process, most of the accounts were correctly labeled. It should be noted that users who shared their credentials with their families and friends living abroad were discarded from our dataset. The reason was that those accounts were not compromised, although their activity matched that of compromised accounts. The accuracy of the classifier would be negatively impacted if those accounts were labeled as benign. In addition, although detection of shared credential accounts is not a goal of UCAAS, we did not count them as false positives if they were flagged by UCAAS, because sharing of credentials is discouraged by the universities.

At UofM, the training data were collected from June 14 to June 28, 2011. The dataset includes 108,366 unique users who had 2,129,275 authentication attempts. After filtering out users who did not have VPN-related activities and conducting the validation process, we got a final training set of 2,441 benign and 87 compromised accounts. The empirical evaluation was done on a different validation set collected from September 14 to October 2, 2011 and consisting of 6,562,153 login sessions from 127,316 unique users. At UIUC, the training data were collected from June 19 to July 2, 2011. There are 104,172 successful logins from 25,530 users in the dataset. After the filtering and validation, we got a final training set of 4,692 benign and 6 compromised accounts. The evaluation set was collected from July 9 to July 23, 2011 and consists of 106,477 logins from 24,979 unique users. The limited number of compromised accounts at UIUC led us to add 10 incidents to the set. Those incidents were detected by the security team during the first half of 2011. We carefully examined the impact of adding those incidents to make sure that they would help to improve model accuracy without affecting the false positive rate.

B. Feature Evaluation

In this subsection, we analyze the effectiveness of the feature set by comparing the proportions of benign and compromised accounts that were flagged by each feature. We also list the coefficients for significant features, as calculated by Weka when building the model. As explained in Section **??**, the coefficients represent the contribution of each feature to the model, so the presence of features with high coefficients indicate accounts that were likely compromised.

The results for the set of suspicious behavior features are shown in Table II. We find that for both institutions, a higher proportion of compromised accounts manifest suspicious behavior. The only exception was that, interestingly, no compromised account was reported as having a suspicious

Institution	Suspicious Behavior	% of benign	% of compromised
UofM	Temporal-Spatial Violation	1.74%	42.08%
	Suspicious IP Addresses	1.67%	0
	Suspicious Usage Pattern	18.03%	37.60%
UIUC	Temporal-Spatial Violation	0.32%	11.88%
	Suspicious IP Addresses	0.24%	2.43%
	Suspicious Usage Pattern	72.97%	75.10%

Table II
PROPORTIONS OF AUTHENTICATION ATTEMPTS FROM BENIGN ACCOUNTS AND COMPROMISED ACCOUNTS FLAGGED AS SUSPICIOUS

Institution	Country	% of benign	% of compromised
UofM	United States	75.67%	19.97%
	Iran	0.64%	38.10%
	China	18.13%	29.97%
	Egypt	0	2.41%
	Japan	0.24%	1.84%
UIUC	United States	79.16%	15.14%
	China	15.66%	68.43%
	Iran	0.23%	10.37%
	Nigeria	0	2.87%
	France	0.36%	1.87%

Table III
DISTRIBUTION OF IP ADDRESS GEOLOCATION

IP address in the training set at UofM. This indicates a drastic evolution of the threat model, since half of the incidents reported in 2010 were linked to suspicious IP addresses. The coefficients for the feature temporal-spatial violation, suspicious IP addresses, and suspicious usage pattern at UofM are 6.81, 8.11 and -35.39, respectively. It is in accordance with our observations that in this model, suspicious IP addresses is highly negatively correlated with compromises while the other two suspicious behaviors are positive indicators of compromises.

We then analyzed the performance of IP-based features. The top five country codes, ASNs, and TLDs are shown in Tables III, IV, and V, respectively. It is interesting to

Institution	ASN	% of benign	% of compromised
UofM	University ASN	37.38%	9.79%
	Comcast ASN	17.06%	3.04%
	4134	3.63%	15.02%
	12880	0	9.20%
	16322	0.02%	6.76%
UIUC	University ASN	25.78%	3.05%
	4134	6.06%	45.08%
	Comcast ASN	22.96%	0.88%
	4812	2.65%	16.18%
	39501	0.06%	10.21%

Table IV
DISTRIBUTION OF ASN

Institution	TLD	% of benign	% of compromised
UofM	.EDU	47.56%	21.79%
	.NET	39.89%	54.73%
	.CN	2.33%	4.82%
	.IR	0.01%	1.21%
	.COM	6.19%	6.94%
UIUC	.EDU	31.70%	15.52%
	.BIZ	0.06%	12.03%
	.NET	49.48%	38.51%
	.CN	3.34%	10.48%
	.FR	0.23%	6.79%

Table V
DISTRIBUTION OF TLDS

Feature	Coefficient	Feature	Coefficient	Feature	Coefficient
US	-0.94	University	-1.25	.EDU	-0.44
IR	1.95	Comcast	-1.16	.NET	0.15
CN	0.50	ASN 1434	0.09	.CN	0.21
EG	8.52	ASN 12880	9.26	.IR	5.02
JP	1.07	ASN 16322	1.35	.COM	0.07

Table VI
DISCRIMINATING COEFFICIENTS OF IP-BASED FEATURES AT UOFM

Institution	Service	% of benign	% of compromised
UofM	Web-based services	46.26%	28.39%
	VPN	53.74%	71.61%
UIUC	Web-based services	4.24%	2.50%
	VPN	95.76%	97.50%

Table VII
SERVICE USAGE DISTRIBUTION

observe the similarities between the two institutions. For instance, the large majority of legitimate activity occurs in the United States and within the borders of the universities, and the top two countries linked to illegitimate activity are similar, along with an ASN of China-Telecom (4134). On the other hand, the rankings of TLDs and ASNs also reveal some differences. Although the training datasets were collected around the same time, we observed that illegitimate activity came from different sets of Internet service providers for the two universities.

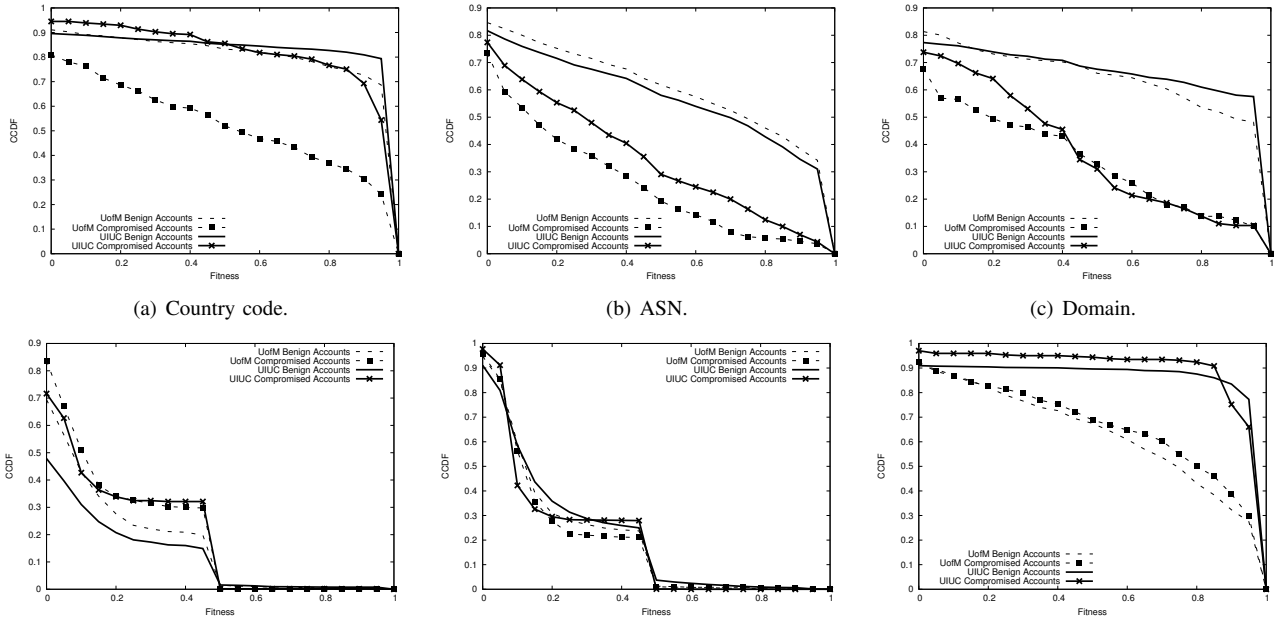
When looking at the discriminating coefficients of some of those features at UofM shown in Table VI, we get the same results. The features of country being China or Iran, ASN being 4134, 12880, 16322, and TLD being .NET, .CN, or .IR all have positive coefficients, while the United States, the University's ASNs, and the TLD .EDU have negative coefficients.

We then reviewed the performance of the resource usage-based features. As shown in Table VII, compromised accounts are more likely to use the VPN rather than web-based services for both universities. Again, the coefficients for the features of web-based service and VPN service at UofM are -0.38 and 0.38, respectively. However, the low proportion of web-based service usage at UIUC prevented us from computing meaningful performance results for the website usage. Thus, Table VIII presents results only for UofM. Once again, the result matches our intuition that a significant proportion of illegitimate activity is linked to library access that has a coefficient of 3.53, while legitimate users accessed the web mail, course portal, and other academic websites more frequently.

Finally, Fig. 3 shows the performance evaluation of profile-based features. The complementary cumulative dis-

Institution	Website	% of benign	% of compromised
UofM	Library	5.83%	44.66%
	Web Mail	34.63%	28.72%
	Course Portal	5.93%	0.78%
	Remote Desktop	1.39%	0
	File Storage	1.22%	0

Table VIII
DISTRIBUTION OF WEBSITES VISITED AT UOFM



(a) Country code. (b) ASN. (c) Domain.

(d) Day of the week. (e) Time of the day. (f) Resource Usage.

Figure 3. Complementary cumulative distribution function of profile fitness for benign and compromised accounts.

tributions of profile *fitness* for the different features indicate that IP-related profile features (country code, ASN, and domain name) are effective in differentiating legitimate from illegitimate activity. However, timing and resource-related features provided poor performance, since the profiles collected for the benign and compromised groups of accounts are indistinguishable.

C. Model Evaluation

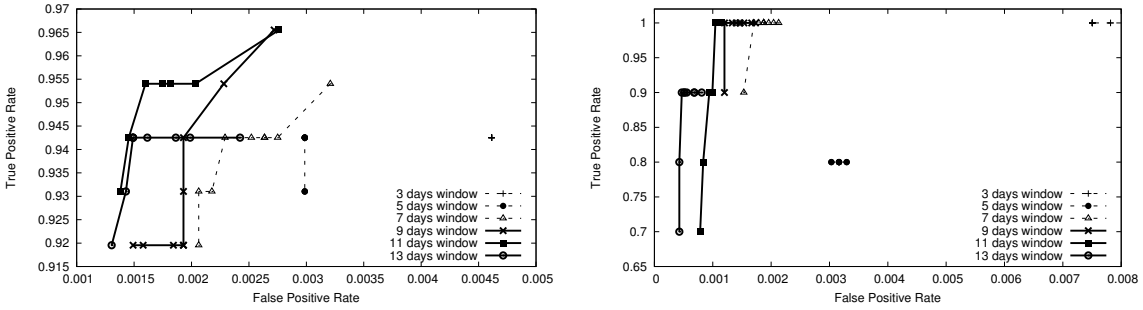
To evaluate the model, we conducted a cross-validation, which is the traditional technique for evaluating machine-learning algorithms. Given the limited number of positives, we use a fivefold cross-validation to ensure an adequate number of positive samples in each set. Since each user may have multiple feature vectors corresponding to multiple days, we grouped the feature vectors by user before partitioning the dataset. By doing so, we prevented artificially good results by removing situations in which different feature vectors were recorded in the training set and the evaluation set for the same user.

An important parameter to choose is the length of the training window. In Fig. 4, we show the detection results collected from five training window sizes. As mentioned before, our goal is to limit the average daily number of false alarms to two. Since UCAAS analyzes about 1,000 unique users per day, we need to achieve a false positive rate (FPR) less than or equal to 0.2%. Under that requirement, the best true positive rates (TPRs) that the system achieved was 95.4% at UofM with a training window of 11 days, and 100% at UIUC with a training window of 9 or 11 days.

D. Empirical Evaluation

We conducted the empirical evaluation by building a model from the training set that was then used to classify the validation set. At UofM, during two weeks starting from September 14, 2011, 126 unique users who never appeared in the training set were flagged. 124 of them were validated by the security team as compromised. The remaining 2 accounts had been shared with family members or friends living in foreign countries. Those results exceeded our expectations, since none of the flagged accounts were false positives. Also, there were no compromised accounts that were detected in other ways but not detected by UCAAS. Therefore, we can conclude that UCAAS achieves better detection recall than any other existing methods in the university.

At UIUC, we conducted the empirical experiment by running UCAAS on the validation dataset collected from July 9 to July 23, 2011. A total of 11 alerts were produced, reflecting 10 accounts already labeled for illegitimate activity in the training set. They appeared here again because those accounts were still compromised when the validation set was collected. The flagged account that was not part of the training set was validated as a true positive. These results are encouraging, because no false positive was generated and a new compromised account was discovered. We also checked with the security team to confirm that the system did not miss a compromised account reported during this period. However, because we collected the validation set at UIUC very close to the time we collected the training set , the lack of newly compromised accounts limits our conclusions regarding the overall accuracy of UCAAS at UIUC. The



(a) UofM. (b) UIUC.

Figure 4. ROC curves.

fact that UIUC has 10 times fewer active accounts than UofM likely plays an important role in explaining the large difference between the numbers of compromised accounts at the two universities. As mentioned in Section II, the difference in the number of active accounts is due to the difference in account expiration rules. Indeed, we found that half of the 124 compromised accounts detected at UofM were alumni or former employee accounts. In conclusion, results at both universities revealed excellent performance. We were surprised to see that at UofM, the model built in June offered perfect accuracy on September-October data, indicating that it remained effective even after a few months.

E. Lessons Learned

We learned that the temporal-spatial violation feature was the main reason for false positives generated during initial testing phases. The incorrect violations were due to three types of events: 1) users connected through more than one VPN client, 2) users accessing campus services via remote desktops, and 3) imprecise geolocation information.

The first two issues were mitigated by the fact that our approach combines complementary sets of features. A temporal-spatial violation has to be associated with additional suspicious behavior captured by other features to raise an alert. To address the last issue, of unreliable geolocation, we revised location-related features to work at the country level rather than the city level, and use a GeoIP database [8] that is accurate enough at the country level. However, we observed that attackers who own credentials from multiple institutions can login to one university account via the VPN of another university. By doing so, they can hide geographic location information and evade our detection method. Therefore, the city-level geolocation information is essential to covering those cases. As future work, we plan to integrate better geolocation lookup approaches [9] to increase the robustness of UCAAS.

V. Related Work

Research efforts that have tried to understand the targets of account theft through botnet takeover [3] or phishing target analysis [2] pointed out that attackers are mostly attracted by financial and payment systems, since their credentials can be directly linked to monetary gains. The main difference with compromised university accounts is that users can easily spot illegitimate transactions, since they have access to the history of account activity. However, compromised university accounts can be used stealthily for months or even years without being noticed by users. Therefore, a centralized compromised accounts assessment system is critical for academic institutions. In [10], the authors analyzed credential-stealing attacks at UIUC based on forensic data collected over five years. The results showed that attackers not only accessed university resources with the compromised credentials, but also harvested additional accounts and resources by exploiting vulnerabilities. They concluded that boundary protections (e.g., firewalls) are insufficient for this threat, and that institutions need sophisticated user action monitoring systems.

Various statistical methods have been successfully applied to detection of fraudulent activities in other security domains. For different applications, the detection tools vary because of the nature of the problems as well as the diversified data types [5]. Applications include credit card fraud detection [11], [12], [13], telecommunications fraud detection [14], [15], [16], and intrusion detection systems [17], [18], [19]. Another interesting security area in which machine-learning methods are heavily used is that of malicious domain and URL detection [20], [21], [22], where features such as lexical or network features are used to distinguish malicious domains and URLs.

Our approach follows the general methods of anomaly and fraud detection, by which we extract a set of features and apply a statistical model to detect suspicious activities. We note that our work addresses the additional challenge of open university environments with very diversified user behavior. The work we found most related to ours is a case study on anomaly detection for VPN [23]. The purpose of the study was to identify suspicious authentication activities through clustering, and geographic distance was used as the main feature. Our work significantly extends that approach by extracting a larger set of features and analyzing them automatically so that only a minimal manual effort is required.

VI. Conclusion

Large academic institutions are exposed to the difficult challenge of protecting user accounts while supporting a

wide set of services with limited security resources. This paper presents the University Credential Abuse Auditing System (UCAAS), a machine-learning approach for automatic detection of account compromises that abuse the VPN service. It considers a large set of automatically generated features. These features are evaluated on their ability to identify illegitimate behavior. A logistic regression classifier is then used to flag accounts that are likely to be compromised. The system was trained and evaluated across two large universities and has been used by the operations team to identify a total of 125 compromised accounts in our two-week trial. Empirical validation shows that UCAAS offers high detection accuracy with no false positives across the two universities. This work is the result of an extensive collaboration, not only between researchers at two different institutions, but also between researchers and security analysts who deal with the issue of account compromise on a daily basis.

ACKNOWLEDGMENT

We wish to thank Paul Howell, who is the chief information technology security officer at UofM, for support in data collection and access, discussion, and result validation. We are also grateful to the security team at UIUC for their help; we particularly wish to thank Michael Corn, Vlad Grigorescu, Warren Raquel, and Bill Gambardella. We would also like to thank Carol Livingstone from the Division of Management Information at UIUC for her support in collecting and analyzing the demographic dataset.

This project has been sponsored at UIUC by the Air Force Research Laboratory (AFRL), and we are thankful for the support of Patrick Hurley. This work was supported at UofM in part by the Department of Homeland Security (DHS) under contract number NBCHC080037, by the National Science Foundation (NSF) under contract numbers CNS 1111699, CNS 091639, CNS 08311174, and CNS 0751116, and by the Department of the Navy under contract N000.14-09-1-1042. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] Trend Micro, "Data-stealing malware on the rise: Solutions to keep businesses and consumers safe," http://us. trendmicro. com/imperia/md/content/us/pdf/threats/securitylibrary/, 2009.

[2] Anti-Phishing Working Group, "Phishing activity trends report," http://www.antiphishing.org/phishReportsArchive.html, 2010.

[3] B. Stone-Gross, M. Cova, L. Cavallaro, B. Gilbert, M. Szydlowski, R. Kemmerer, C. Kruegel, and G. Vigna, "Your botnet is my botnet: Analysis of a botnet takeover," in *Proc. of the 16th ACM Conference on Computer and Communications Security*, 2009, pp. 635–647.

[4] J. R. Young, "Academic publisher steps up efforts to stop piracy of its online products," http://chronicle.com/article/Academic-Publisher-Steps-Up/128031, 2011.

[5] R. J. Bolton and D. J. H, "Statistical fraud detection: A review," *Statistical Science*, vol. 17, pp. 235–249, 2002.

[6] "Weka 3 - Data Mining with Open Source Machine Learning Software," http://www.cs.waikato.ac.nz/ml/weka/.

[7] J. G. Orme and T. Combs-Orme, *Multiple Regression with Discrete Dependent Variables*. Oxford University Press, 2009.

[8] *MaxMind*. http://www.maxmind.com/app/ip-location, 2011.

[9] Y. Wang, D. Burgener, M. Flores, A. Kuzmanovic, and C. Huang, "Towards street-level client-independent IP geolocation," in *Proc. of the 8th USENIX Symposium on Networked Systems Design and Implementation*, Mar. 2011.

[10] A. Sharma, Z. Kalbarczyk, R. Iyer, and J. Barlow, "Analysis of credential stealing attacks in an open networked environment," in *Proc. of the Fourth International Conference on Network and System Security*. Washington, DC, USA: IEEE Computer Society, 2010, pp. 144–151.

[11] E. Aleskerov, B. Freisleben, and B. Rao, "Cardwatch: A neural network based database mining system for credit card fraud detection," *Proc. of the IEEE IAFE 1997 conference on Computational Intelligence for Financial Engineering*, pp. 220–226, 1997.

[12] S. Panigrahi, A. Kundu, S. Sural, and A. K. Majumdar, "Credit card fraud detection: A fusion approach using Dempster-Shafer theory and Bayesian learning," *Inf. Fusion*, vol. 10, pp. 354–363, Oct. 2009.

[13] R. Chen, T. Chen, Y. Chien, and Y. Yang, "Novel questionnaire-responded transaction approach with SVM for credit card fraud detection," in *Proc. of the 2nd International Conference on Advances in Neural Networks*, vol. 2, 2005, pp. 916–921.

[14] D. Agarwal, "An empirical Bayes approach to detect anomalies in dynamic multidimensional arrays," in *Proc. of the Fifth IEEE International Conference on Data Mining*. IEEE Computer Society, 2005, pp. 26–33.

[15] K. C. Cox, S. G. Eick, G. J. Wills, and R. J. Brachman, "Visual data mining: Recognizing telephone calling fraud," *Data Mining and Knowledge Discovery*, vol. 1, pp. 225–231, 1997.

[16] C. Phua, D. Alahakoon, and V. Lee, "Minority report in fraud detection: Classification of skewed data," *SIGKDD Explor. Newsl.*, vol. 6, no. 1, pp. 50–59, 2004.

[17] F. Esponda, S. Forrest, and P. Helman, "A formal framework for positive and negative detection schemes," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 34, no. 1, pp. 357–373, 2004.

[18] K. A. Heller, K. M. Svore, A. D. Keromytis, and S. J. Stolfo, "One class support vector machines for detecting anomalous windows registry accesses," in *Proc. of the Workshop on Data Mining for Computer Security*, 2003.

[19] J. D. Brutlag, "Aberrant behavior detection in time series for network monitoring," *Proc. of the 14th Systems Administration Conference (LISA 2000)*, Dec. 2000.

[20] C. Whittaker, B. Ryner, and M. Nazif, "Large-scale automatic classification of phishing pages," in *Proc. of 17th Annual Network and Distributed System Security Symposium*, 2010.

[21] M. Antonakakis, R. Perdisci, W. Lee, V. Nikolaos, and D. Dagon, "Detecting malware domains at the upper DNS hierarchy," pp. 1–16, 2011.

[22] J. Ma, L. K. Saul, S. Savage, and G. M. Voelker, "Beyond blacklists: Learning to detect malicious web sites from suspicious URLs," in *Proc. of the SIGKDD Conference.*, 2009.

[23] M. Chapple, N. Chawla, and A. Striegel, "Authentication anomaly detection: A case study on a virtual private network," in *Proc. of the 3rd Annual ACM Workshop on Mining Network Data*. New York, NY, USA: ACM, 2007, pp. 17–22.