Learning from Early Attempts to Measure Information Security Performance

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 Computer Science and Engineering, 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

The rapid evolution of threat ecosystems and the shifting focus of adversarial actions complicate efforts to assure security of an organization's computer networks. Efforts to build a rigorous science of security, one consisting of sound and reproducible empirical evaluations, start with measures of these threats, their impacts, and the factors that influence both attackers and victims. In this study, we present a careful examination of the issue of account compromise at two large academic institutions. In particular, we evaluate different hypotheses that capture common perceptions about factors influencing victims (e.g., demographics, location, behavior) and about the effectiveness of mitigation efforts (e.g., policy, education). While we present specific and sometimes surprising results of this analysis at our institutions, our goal is to highlight the need for similar in-depth studies elsewhere.

1 Introduction

Institutions and their users are the target of a variety of threats including malware, botnets, Distributed Denial of Service (DDoS), identity theft, and spam, which challenge the availability of networks and the confidentiality of users' data. To understand how those threats succeed in impacting the security of institutions and how they can be mitigated, researchers must identify and measure the factors that influence both attackers and victims. Quantifying security is particularly important if we are to move from incremental improvements that simply keep pace with those evolving threats to structural or fundamental changes to the security landscape.

While a broad range of security analyses have been attempted—including analyses of what types of rogue software are installed on compromised machines [11], what data are collected by attackers [14, 30], what data are sold [13], and how exploits lead to financial gains [17]—there remains an important gap between the results collected and the development of meaningful security measurements that support the decisions that need to be made to protect systems and networks.

This paper presents our attempt to bridge this gap in the context of account compromise at academic institutions. Our goal has been to empirically evaluate common perceptions about the factors influencing account compromise victims. We translated those perceptions into hypotheses and conducted a study of account compromise incidents at two large university environments over multiple years to confirm or deny our hypotheses. The focuses of our hypotheses are on victim information (e.g., demographics, location, behavior) and on the effectiveness of several attempts to implement proactive controls (e.g., policy, education) over the security landscape at our institutions. Academic institutions are particularly interesting for this type of study because they offer a significant degree of visibility into both their unique threat landscape (i.e., targeted attacks) and the vulnerability surfaces of their infrastructure. Moreover, they can implement, on reasonably short timescales, proactive and reactive measures to improve their security posture.

While our study offers several interesting preliminary positive findings (e.g., education level, security training, and network topological location were significant factors in the susceptibility of victims) and negative findings (e.g., neither gender nor geographic location impacts susceptibility), we make no claim that these results generalize beyond our institutions, nor that we have exhaustively explored or proven any properties about the organizational perspective. Rather, our goal is to highlight this perspective and carefully explain our empirical methodology to encourage other organizations and the community at large to pursue a more quantified approach to security analysis.

2 Background

Starting in late 2010, we engaged in a pilot project with the Office of Information and Infrastructure Assurance

(IIA) at the University of Michigan (UofM) and the security team at the University of Illinois at Urbana-Champaign (UIUC) to investigate metrics for enterprise security. Here our goal was similar to that of many existing security metrics efforts:

> "IT security metrics provide a practical approach to measuring information security. Evaluating security at the system level, IT security metrics are tools that facilitate decision making and accountability through collection, analysis, and reporting of relevant performance data. Based on IT security performance goals and objectives, IT security metrics are quantifiable, feasible to measure, and repeatable. They provide relevant trends over time and are useful in tracking performance and directing resources to initiate performance improvement actions." [20]

As a first step in this process, the security teams identified account compromise as a pressing security problem, as well as an area in which existing data could be analyzed to build useful metrics.

Figure 1: Online reselling of stolen credentials.

Most universities use basic ID and password authentication methods across systems; therefore, once account credentials are discovered by an illegitimate entity (a person or an automated agent), the account becomes fully compromised. Others have shown that adversaries usually steal university account credentials by attacking authentication mechanisms (e.g., by guessing passwords, exploiting vulnerabilities, or installing trojans), phishing, or social engineering [26].

Such compromised accounts are useful for a wide variety of reasons. Some traditional exploits include resource abuse, accessing of confidential information, spamming, and further credential harvesting. More interestingly, we find that attackers also seek monetary gains by reselling the stolen credentials (as shown in Figure 1). As a motivation for our work on VPN abuse detection [35], an in-depth analysis of the malicious activities revealed that the motivation of the attackers is to gain free and unfettered access to information. We showed the stolen credentials were used primarily to download scholarly publications [34] and circumvent state-sponsored censorship (i.e., in China and Iran).

To mitigate those threats, both *proactive methods* and *reactive procedures* are deployed. While reactive procedures, including compromised account detection and password resetting, are aimed at minimizing the damage after compromises happen, proactive methods are used to reduce the likelihood that end users' accounts become compromised. The proactive methods adopted at universities are mostly part of *education* and *policy* efforts. Education includes publication of online materials [5] and offering of workshops and security quizzes. Preventive policies include password creation and update policies [1, 4], as well as the implementation of automated account locking after several unsuccessful login attempts.

3 Related Work

Online account security is an important issue that has plagued institutions for a long time. Hackers can breach institutions' networks and gain access to confidential information [10]. Also, they can use malware to steal and harvest user credentials. As shown in [30], the Torpig botnet, a sophisticated malware program designed to steal sensitive information, collected 8,310 account credentials from 410 different institutions over a period of ten days. Another emerging threat facing Internet users is phishing techniques: a real-world phishing experiment [19] revealed that 52.3% of participants clicked on phishing emails and 40.1% of participants gave their information to the phishing sites.

Extensive research has been done to study how best to safeguard online accounts. The efforts to understand the demographics of phishing victims found that women were more likely to fall for phishing attacks [16, 27, 19] and participants aged 18-25 years old were more vulnerable than other participants [18, 27]. Research in security user education have revealed different ways to improve the ability of users to protect their information against phishing, including online materials [3, 6, 2], contextual training [16], and online games [28]. The studies tested the effectiveness of those methods, and they found both that online materials improved user ability drastically if users read them carefully [19], and that education training reduced the tendency to disclose information disclosure by 40% [27].

In addition, some research projects focused of the issue of password security, which is claimed to be one of the weakest parts of secure systems. An analysis on a large set of real-world passwords [31] found that current password policy is not always effective, because a subset of users pick passwords that can easily be cracked

despite complying with the password policy. Research also pointed out that the quick proliferation of password-protected sites inspires credential reuse among multiple accounts, which in turn jeopardizes the security of all the systems [15]. A poll survey [8] reported that a third of 676 Internet users admitted they reused online credentials. In addition, a recent study [32] showed that users tend to choose passwords that are easy to break when they use fingerprint authentication and, as a result, the strength of multi-level authentication systems is weaken.

4 Methodology

4.1 Data Used in this Study

During 2009, 2010, and the first six months of 2011, IIA recorded 6,600 abuse-related trouble tickets; 1,200 of these were security incidents. The tickets contain information including ticket creation time, category of incidents, comments of security operators, and victim responses. Those tickets are "self-reported" in that the related incidents come directly from university departments, organizations, and service groups, as well as UofM students, staff, and alumni. The tickets cover a wide range of incidents, including "unauthorized exposure of private personal information, computer break-ins and other unauthorized use of UofM systems or data, unauthorized changes to computers or software, equipment theft or loss, and interference with the intended use of information technology resource" [7]. By far the most dominant group of tickets involve unauthorized use of UofM systems, with 822 incidents confirmed (via IIA staff) over two and a half years. At UIUC the number of tickets related to malicious account activity was 178 in 2011. In addition to the trouble tickets, data used in our study include the following:

- *Authentication logs:* The authentication systems deployed at UofM and UIUC share a similar structure. They are both based on Kerberos with different authentication portals (e.g., Web-based services, wireless, VPN). The authentication logs provide useful timing, location, and service usage information, which are critical for tracking user activities.
- *Victim information:* Victim demographics include gender, age, role, nationality, appointment, education level, and working address.
- *Effectiveness of some mitigation strategies:* Records of previous actions that aimed to prevent user account compromise were collected. For example, these include the results of scanning for weak passwords and the list of users who passed a computer security quiz.

University/Year		Total Population	Victims	Victim/Total
UofM	**2009**	75,992	136	0.18%
	2010	76,944	279	0.36%
	2011	74,346	222	0.30%
UIUC	**2011**	54,612	178	0.32%

Table 1: Datasets for demographic analysis.

Group	Variable	Type	Details
Student	Gender	Binary	Male, Female
	Age	Categorical	<19, 20-21, 22-23, 24-25, 26-30, 31-35, >35
	Education	Categorical	Undergraduate, Graduate, Others
	Citizenship	Binary	U.S. Citizen, Non-U.S. Citizen
	Department	Categorical	
Faculty/ Staff	Gender	Binary	Male, Female
	Age	Categorical	<30, 30-39, 40-49, 50-59, 60-64, ≥65
	Education	Categorical	2-year degree, Bachelor (equal), Masters', PhD and above, Specialist, Others
	Citizenship	Binary	U.S. Citizen Non-U.S. Citizen

Table 2: Demographic variables explored.

A brief note on the sensitive nature of the data: the main risk of this analysis, as identified by the researchers, is informational risk—that is, the psychological, legal, social, and economic damage resulting from inappropriate use or disclosure of information. While the security teams at both institutions oversaw compliance with university practice for the projects, additional steps were taken to mitigate these harms, including assurances that access was for valid statistical purposes, that the researchers used the data appropriately and followed procedures, in some cases that the data were made inherently non-disclosive of any sensitive information, that technical controls surrounding access prevent the unauthorized removal of data, and that the results produced did not contain any protected information [23].

4.2 Demographic Factors

Table 1 shows the data sets used in our demographic analysis. At UofM, IIA collected demographic data for the total on-campus population from January 2009 to July 2011. Over this period, there were 637 victims of account compromises, excluding alumni and former employees. Similar demographic data were collected at UIUC for the year of 2011, and a total of 178 incidents were recorded during this period. We separated accounts into two populations: students and faculty/staff. Within each group, we considered a variety of factors, including gender, age, education level, and citizenship. Table 2 lists this breakdown and the associated subcategories.

Methodology In our analysis, we use *logistic regression*, a form of statistical multiple regression models [21],

University/Year		% of total pop.	% of victims
UofM	2009	45.16%	57.35%
	2010	45.51%	62.36%
	2011	46.28%	65.17%
UIUC	2011	19.68%	26.97%

Table 3: Proportion of faculty and staff members in the total population and in the victim population.

in order to understand and explain the relationship between multiple user demographic factors and the users' susceptibility to compromises. The multiple regression allows us to predict the possibility of compromise based on one demographic factors, *holding other factors constant*. For example, this technique is useful to test whether undergraduate students are more likely to become victims than graduate students who are of the same gender, age, citizenship, and department. Thus, we can determine which demographics are predictors of compromise.

The estimated regression model is:

$$L = a + \sum B_i X_i.$$

L in the equation represents the natural logarithm of the odds that the event represented by the dependent variable, which in our case is the account compromise, happens. When $\hat{p}$ represents the estimated probability that the event happens,

$$L = \ln \frac{\hat{p}}{1-\hat{p}}.$$

For each predicting variable X_i, the coefficient B_i indicates the amount of change in the natural logarithm with one unit of change in the predictor variable, adjusting for the other variables. For each predictor variable, we test the *null hypothesis* $H_o : B_i = 0$ against the alternative $H_a : B_i \neq 0$. If the null hypothesis is valid, we can conclude that there is not sufficient evidence to indicate that variable X_i is significant in predicting susceptibility, *holding other variables constant*. We use *p-value* as the test statistic and test at a 95% confidence level ($\alpha = 0.05$). The predicting variables with $p - value < \alpha$ are considered significant.

In addition, we recognize that there might be a multicollinearity problem (i.e., correlation among different predictor variables), which could create the false impression that a predictor was significant [21]. Therefore, we use the *Variance Inflation Factor (VIF)* as the diagnostic statistic for multicollinearity. The cutoff value of VIF for determining the presence of multicollinearity is usually 10, which means that values of VIF exceeding 10 are regarded as indicating multicollinearity [12]. By applying this diagnostic on our dataset, we found there is no multicollinearity issue.

Comparing Students with Faculty/Staff One of the questions we wished to explore was *(Q1): which sub-*

Factor	University/Year		p-value	coefficient
Undergraduate	UofM	2009	***0.009***	2.957
		2010	***<0.001***	3.520
		2011	***0.020***	3.489
	UIUC	2011	0.958	-10.733
Age (20-21)	UofM	2009	***0.002***	1.219
		2010	***0.004***	0.823
		2011	***0.017***	0.896
	UIUC	2011	0.410	-0.472
Citizenship	UofM	2009	0.520	0.315
		2010	0.659	-0.126
		2011	0.128	-0.460
	UIUC	2011	***0.007***	0.5433

Table 4: Significant influential variables in students.

population is more likely to be infected? The proportion of faculty and staff members in the total population and in the victim population are shown in Table 3. We note that there are more faculty and staff members at UofM, because employees of the university's hospital are part of our dataset. It can be seen that faculty and staff members show a higher proportion of victims than students do at both universities. In the statistical analysis, the role of faculty/staff also impacts the susceptibility significantly, with p-values of 0.0017 and 0.0006 at UofM and UIUC respectively. In contradiction to our initial belief, members of the faculty/staff population are more likely to become victims than members of the student population are.

Student Group For the student group, we applied logistic regression on all the datasets, and once again we extracted only significant factors using a *p-value* below 0.05. Here we were interested in: *what roles gender (Q2), age (Q3), education-level (Q4), citizenship (Q5), and department (Q6) play in the compromise of student accounts?*

Table 4 shows that *education-level* and *age between 20 and 21 years* appear as significant factors influencing susceptibility at UofM. Undergraduate students, who represent 64.78% of the total student population, constituted 87.36% of the student victims in 2011. With p-values lower than 0.05, undergraduate students appear more likely to be compromised than graduate students. Students 20 to 21 years old are the most susceptible to becoming compromised at UofM. About 28.5% of the total student population falls into this age group, but they accounted for 38.89%, 37.14%, and 47.13% in the student victim population in 2009, 2010, and 2011 (with p-values of 0.002, 0.004, and 0.017). However, there is insufficient evidence to show that those two factors are also significant at UIUC.

Citizenship appears as a significant predictor of susceptibility for UIUC but not for UofM. In 2011, foreign students accounted for about 22% of total students at UIUC. Meanwhile, 33.58% of the student victims were foreign, indicating that foreign students were more sus-

Factor	University/Year		p-value	coefficient
Citizenship	**UofM**	**2009**	***0.001***	1.307
		2010	***< 0.001***	2.195
		2011	***< 0.001***	2.292
	UIUC	**2011**	0.414	0.368
Age (<30)	**UofM**	**2009**	***< 0.001***	1.868
		2010	***< 0.001***	1.846
		2011	***< 0.001***	2.223
	UIUC	**2011**	0.700	-0.487
Age ($\geq$ 65)	**UofM**	**2009**	0.050	0.833
		2010	***0.021***	0.677
		2011	***0.007***	0.987
	UIUC	**2011**	0.958	0.063

Table 5: Significant influential variables for faculty/staff.

ceptible to compromise than domestic students were. However, similar results were not found at UofM. In addition, we found that *gender*, which has a p-value of 0.07, is potentially a useful predicting variable at UIUC, where males are more susceptible than females. The data show that 64.92% of the student victims were male, while only 53.84% of the total student population was male.

Faculty/Staff Group We performed a similar analysis on the faculty/staff group. *What roles do gender (Q7), age (Q8), education (Q9), and citizenship (Q10) play in account compromises?* Table 5 shows the significant factors that influence victims among faculty and staff members. Results reveal that *citizenship* and *age* are significant in predicting user susceptibility to account compromise at UofM. Foreign employees are more likely to become victims than U.S. citizens are. While about 2.1% of employees at UofM are not U.S. citizens, they account for 8%, 17.18%, and 18.87% of the faculty/staff victims in 2009, 2010, and 2011 respectively (with p-values lower than or equal to .001). Although we observed more foreign faculty and staff members than domestic employees compromised, the null hypothesis test shows insufficient evidence to conclude that citizenship is a significant predictor at UIUC.

4.3 Temporal Factors

We are also interested in understanding when account compromises occur. In particular, we wish to know *(Q11): whether the incidence of compromises varies at different times of the year.* We performed a time series data analysis on the monthly number of tickets at UofM via the "Holt-Winters" exponential smoothing procedure [33]. The analysis decomposes the time series data into *long-term trend* and *seasonality*, and while the long-term trend shows the number of incidents increasing, the fit to the seasonality is poor. As shown in Figure 2, the predicted seasonal effects are absent from the real-world observation. Also, an issue to notice when evaluating such temporal effects is that the timestamps linked to tickets represent ticket creation time rather than compromise

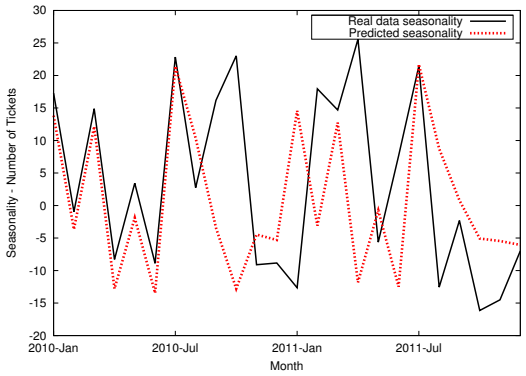


Figure 2: Seasonality in the number of tickets at UofM.

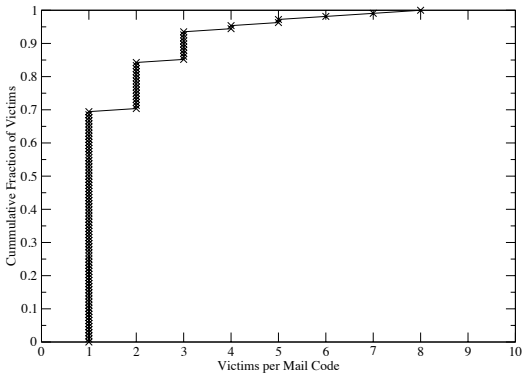


Figure 3: Cumulative fraction of victims per location.

time. As a result, they may be too coarse-grained for a monthly frequency analysis.

4.4 Geographical Factors

In this analysis, we attempted to determine *(Q12): whether victims are clustered geographically.* For example, are specific buildings or campuses more susceptible than others? We hypothesized that attackers taking advantage of user behavior or targeting shared infrastructure will likely have an impact on a victim's susceptibility. Figure 3 shows the breakdown of victim university locations by university mail code at UofM. Mail codes offer a coarse-grained geographic measure to group university buildings. As one can see, nearly 70% of mail codes contained only one victim, with no mail code containing more than eight victims over the two-year period.

4.5 Topological Factors

While the above analysis showed no strong geographic bias, we further hypothesized that infrastructure vulner-

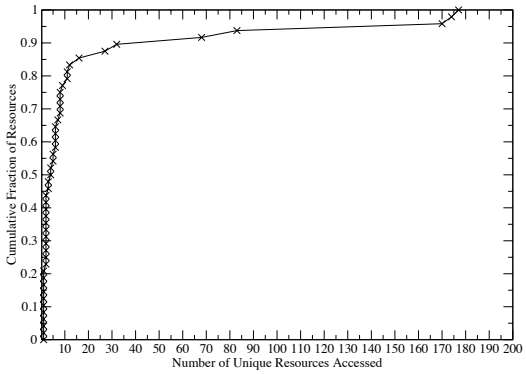


Figure 4: The cumulative number of unique resources used by the victims in each administrative domain.

abilities could be an important reason for compromises. Here we modify our question: *(Q13) Are victims clustered topologically?* We are concerned not with physical location, but with virtual location in the network topology. Figure 4 shows the victim locations, as defined by resources used in a given network administrative domain. A handful of networks stand out as clusters of activity. The top five include two computer laboratory domains, the VPN domains, one school building, and the electronic library resources.

4.6 Service Behavior

At both universities, various online services are provided, among which three basic services occupy the majority of user activities: Web-based services, wireless service and VPN service. Each service authenticates separately, and therefore we can infer usage frequency of different classes of service. We are interested in *(Q14) Whether usage patterns of victims, attackers, and benign users vary.*

We evaluate the portal usage model of three different groups of users at UofM: users who have never been reported as behaving maliciously (referred to as *normal users*), victims who were under the control of adversaries (referred to as *victims/adversaries*), and previous victims who have been compromised but now claim to be benign (referred to as *previous victims*). Since we aim to find whether a compromise is related to specific services, the comparison between normal users, previous victims and victims/adversaries might suggest a good way to design a compromised account detection solution or to enhance a university policy. We calculated the usage percentage of each service at UofM based on the number of successful login sessions during a period of one week in February 2011 at UofM. The data include 113,644 normal users, 261 previous victims and 121 victims/adversaries. We, in Figure 5, show the complementary cumulative percentage of users by the proportion of their usage of each portal. Curves to the upper left indicate populations who use this service frequently and curves to the lower right indicate populations of infrequent users. In examining the figures, we see that previous victims were heavy users of Web services, while normal users and victims/adversaries had similar profiles. Victims/adversaries and previous victims shared similarities in wireless usage, while normal users were much heavier users of the service. Normal users and previous victims were nearly nonexistent in VPN usage, while Victims/adversaries showed heavy usage.

	# of total	# of compromised	Pr (compromise)
Total population	550,000	380	0.069%
Weak password	2,284	12	0.525%

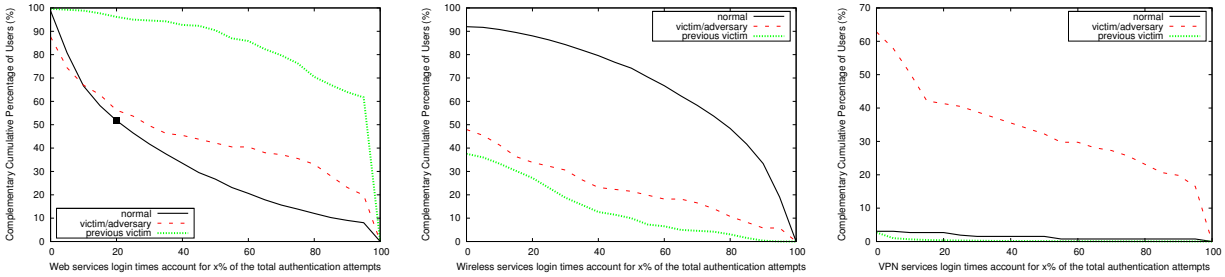
Table 6: Probability of account compromise for the total population and the weak-password accounts.

4.7 Password Strength

Password creation policies aim to prevent the adoption of weak passwords that could easily be cracked. At UofM, the operation team performs weak password scanning every six months, in an effort to reduce the impact of weak passwords. Here, we explore the question of *(Q15) whether accounts with weak passwords more likely to be compromised.*

At UofM, 380 accounts were compromised, while 2,284 accounts with weak passwords were detected in 2012. Only 12 accounts are present in both sets, accounting for 3.16% of compromised accounts and 0.52% of weak-password accounts. Given the small intersection, we may infer that while weak passwords do play some role in the account compromises, they are a very minor attack vector.

However, as shown in Table 6, the probability of a weak-password account being compromised is much higher than the probability of compromise for the total population. Note that the total number of accounts at UofM is 550,000 (the figure is high because alumni accounts can continue to use some university services). In order to determine the significance of this observation, we performed a *test of homogeneity* [9], which is a statistical hypothesis testing method for categorical data. Here, the null hypothesis, H_o, is that accounts with weak passwords have the same probability of compromise as other accounts. The alternative hypothesis, H_a, is that weak-password accounts have a higher probability of compromise. We would reject H_o if $p - value < 0.5$. In our study, the test result has a deviance of 28.09 and a p-value of 1.16^{-16}. Therefore, we reject H_o and conclude that there is sufficient evidence to show that accounts with weak passwords are more likely to be compromised.



(a) Web-based service. (b) Wireless service. (c) VPN service.

Figure 5: Complementary cumulative percentages of users whose usage of the service is more than $x\%$ of the total usage. For example, the square point in (a) indicates that for about 50% of normal users, the web-based service usage accounts for more than 20% of their total service usage (i.e., $\frac{number\ of\ login\ sessions\ to\ web-based\ services}{total\ authentication\ attempts} > 20\%$).

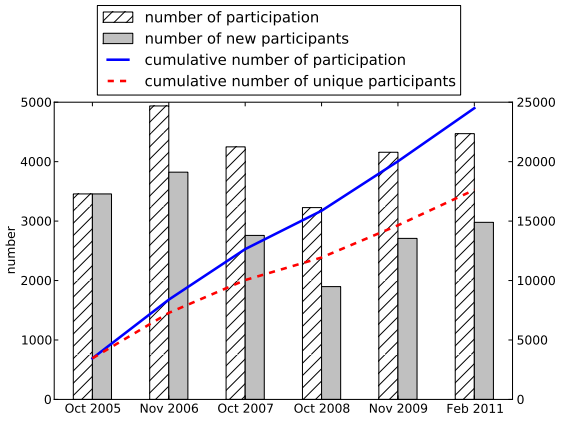


Figure 6: Participation in security quiz.

4.8 Education Impact

Since 2005, students at UofM have participated in an annual voluntary online computer security quiz. We were curious to know *(Q16) whether taking the security quiz positively impacts users' ability to keep accounts secure.* We examined the list of people who passed the quiz (referred to as *educated users*) and found 17,625 distinct users. Among those users, 23 were victims of account compromise. Figure 6 shows the number of participants by year. The quiz can be taken multiple times, and new users who have not taken the quiz before represent on average 71.94% of the total number of users. 27.53% of the participants took the quiz over multiple years, or several times during a single year.

To understand the effectiveness of education, we compared the probability of compromise for two groups in 2010: the educated user group and the total on-campus student group. The total on-campus student population represents 41,924 students, among which 9,227 had passed the security quiz. 105 student accounts were reported to be compromised in 2010; among them nine were from the educated user population. The probability of educated users becoming compromised was 0.1%, while the total on-campus student population had a probability of 0.25%. Again, we applied a test of homogeneity to determine whether this difference is significant. Here, the null hypothesis, H_o, is that all the students who has passed the security quiz have the same probability of becoming victims as those who have not, while the alternative, H_a, is that students who pass the quiz are less likely to be compromised. The results of the test show a deviance of 13.52 and a p-value of 2.36^{-4}. Since the p-value is less than 0.05, we can reject H_o and conclude that people who passed the quiz were less likely to become victims.

However, we can only infer that people who took and passed the security quiz had better awareness and knowledge of security. Thus the question of how effective the security quiz is in mitigating the compromise of accounts remains unanswered in our work.

5 Lessons Learned

We believe in the need for relevant measurements and metrics as we seek to build a science of security. As a result, one effort of this work has been to think beyond the tactical response to our analysis and answer more fundamental problems, such as those of choosing appropriate metrics, creating methods for validating metrics, comparing methods for metric computation and collection, understanding the appropriate uses of metrics (operations, evaluation, risk management, decision making), and building the ability to measure operational security values [25].

From Observations to Hypotheses A striking lesson from this work is the reminder that human observation, while being the launching point for scientific inquiry, is itself only the beginning of the process. Several of

the hypotheses we developed based on our observations proved to be invalid following thorough statistical testing. For example, with regard to Q1, we expected that less aversion to risks and lower levels of experience would make students more susceptible to compromise than the faculty/staff group. With respect to Q2, previous studies had shown that gender played a role in phishing susceptibility [16, 27, 19], and we were interested in determining whether this factor was also influential in our datasets. The finding that either it did not matter or male students were more susceptive(in the case of UIUC) was unexpected. Answers to Q11 also ran counter both to our expectation and to the security teams' experiences. Observations, such as these relating to on-campus student populations or overall traffic volumes, exhibited strong seasonal patterns per semester. The lack of those patterns in our data was surprising. Finally, we hypothesized that Q12 would be an excellent metric, as we felt geographic location encompassed factors, such as targeted and shared environments, that could influence subpopulation susceptibility. It turned out not to be the case.

Publication and Validation An important challenge in a variety of security domains is the design of reproducible experiments [29, 24, 22]. Poor metric selection, misapplication of scientific analysis techniques, or inadequate technical details will greatly diminish the value of empirical results. We experienced this most acutely in the use of trouble tickets as the main source of metrics in our study. Uncertainty in the date of compromise, the source of trouble tickets, or the ticket reporting methodology (e.g., a large break-in could have been recorded in a single or multiple ticket?) challenge the validity of our results.

Generalization In addition to building confidence in the findings of an individual study, cross-validation of those findings helps us to understand their generalizability. For example, we found that the lack of agreement across institutions for questions Q2—Q10 was interesting. While we believe differences between UofM and UIUC metrics and evaluations may exist, we hypothesize that the relevant social, behavioral, and economic factors do not have as consistent as impact on compromises as one might assume.

Facilitating Decision-Making Metrics are, of course, just the first step in building more secure enterprises. Using metrics to answer important questions and guide future actions is necessary next step in any formal security process. In our analysis, this was best highlighted in the answers for Q15—Q16, which confirmed the role of education and password strength in account compromise. However, such correlations do not help us make value judgements about the utility of education or password-cracking efforts. Are the limited number of potentially impacted accounts worth the effort? How do these efforts compare with other options?

6 Conclusion and Future Work

In this paper we presented some of our preliminary efforts in trying to understand the security of academic enterprise networks. Focusing on university account compromises, weidentified a series of questions to help guide the general development of security metrics within those networks. We examined those questions in the specific context of account compromise incidents at UofM and UIUC over a three-year period. The questions studied a variety of analysis, including demographic, temporal, geographic, topological, and behavioral factors that may influence compromise, as well as the effectiveness of existing policy and security training efforts at UofM.

As part of future work, we are interested in evaluating our findings as a guide to the future proactive mitigation efforts. Such evaluation we require further collaboration with the operational teams to enhance and evaluate proactive initiatives as well as to investigate more deeply the mod operandi of attackers (e.g., by using honeypots) and the behavior of victims (e.g., by surveying the campus population). We are also aware that our results remain preliminary and may be specific to our two institutions. Therefore, we are interested in involving additional organizations and extending our analysis to a larger scale.

From a broader point of view, our work fits into a long-term effort by the community to build a science of enterprise security that provides a quantified and continuous security process [25]. A key step in reaching that ambitious goal will be to define a framework for interdisciplinary and cross-institutional research on enterprise security. While the challenges are well-known and significant (they include data availability, data sharing, and privacy issues, as well as the need to enable collaboration among stakeholders from different backgrounds and heterogeneous institutions), this work offers a concrete example to guide future efforts.

Acknowledgements

We would like to express our gratitude to all those who contributed to our work. We wish to thank Paul Howell, who is the chief information technology security officer at the University of Michigan, for his support in data collection and access, discussion, and result validation. We are also grateful the security team at the University of Illinois for their help and fruitful discussions; in particular, we wish 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 the University of Illinois for her support of our collection and analysis of the demographic dataset.

This project has been sponsored at UIUC by the Air Force Research Laboratory (AFRL), and we are grateful 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; the National Science Foundation (NSF) under contract numbers CNS 1111699, CNS 091639, CNS 08311174, and CNS 0751116; and the Department of the Navy under contract N000.14-09-1-1042. This material was based in part on work supported by the National Science Foundation, while working at the Foundation. 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] *Choosing and Changing a Secure UMICH Password*.

[2] *PayPal: can you spot phishing?* `http://www.paypal.com/fightphishing`. Accessed in 2012.

[3] *Protect yourself against phishing. The Facebook blog*. `http://www.facebook.com/blog.php?post=81474932130`. Accessed in 2012.

[4] *Requirements for Acceptable Passwords*. `http://www.cites.illinois.edu/passwords/requirements.html`.

[5] *Safe Computing - Information and Technology Services, University of Michigan*.

[6] *Avoid 'phishing' scams. Twitter blog*. `http://blog.twitter.com/2010/02/avoid-phishing-scams.html`, 2012.

[7] IT security incidents. `http://www.safecomputing.umich.edu/main/incident_report.html`, 2012.

[8] *Security at risk as one third of surfers admit they use the same password for all websites, Sophos reports* . `http://www.sophos.com/en-us/press-office/press-releases/2009/03/password-security.aspx`, 2012.

[9] AGRESTI, A. *Categorical Data Analysis*, 2nd ed. Wiley Series in Probability and Statistics. Wiley-Interscience, 2002.

[10] ARTHUR, C. *PlayStation Network: Hackers claim to have 2.2M credit cards*. `http://www.guardian.co.uk/technology/blog/2011/apr/29/playstation-network-hackers-credit-cards`, 2012.

[11] CABALLERO, J., GRIER, C., KREIBICH, C., AND PAXSON, V. Measuring pay-per-install: The commoditization of malware distribution. In *Proceedings of the 20th USENIX Security Symposium* (2011), pp. 187–202.

[12] FARAWAY, J. J. *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*, 1 ed. 2005.

[13] FRANKLIN, J., PAXSON, V., PERRIG, A., AND SAVAGE, S. An inquiry into the nature and causes of the wealth of internet miscreants. In *Proceedings of the 14th ACM Conference on Computer and Communications Security* (2007), pp. 375–388.

[14] HOLZ, T., ENGELBERTH, M., AND FREILING, F. Learning more about the underground economy: A case-study of keyloggers and dropzones. In *Proceedings of the 14th European Conference on Research in Computer Security* (Berlin, Heidelberg, 2009), pp. 1–18.

[15] IVES, B., WALSH, K. R., AND SCHNEIDER, H. The domino effect of password reuse. *Commun. ACM 47*, 4 (Apr. 2004), 75–78.

[16] JAGATIC, T., JOHNSON, N., JAKOBSSON, M., AND MENCZER, F. Social phishing. *Communications of the ACM 50*, 10 (October 2007), 94–100.

[17] KANICH, C., KREIBICH, C., LEVCHENKO, K., ENRIGHT, B., VOELKER, G. M., PAXSON, V., AND SAVAGE, S. Spamalytics: an empirical analysis of spam marketing conversion. In *Proceedings of the ACM Conference on Computer and Communications Security, CCS'08* (Oct. 2008), pp. 3–14.

[18] KUMARAGURU, P., CRANSHAW, J., ACQUISTI, A., CRANOR, L., HONG, J., BLAIR, M. A., AND PHAM, T. School of phish: a real-world evaluation of anti-phishing training. In *Proceedings of the 5th Symposium on Usable Privacy and Security, SOUPS '09* (New York, NY, USA, 2009), pp. 1–12.

[19] KUMARAGURU, P., SHENG, S., ACQUISTI, A., CRANOR, L. F., AND HONG, J. Teaching Johnny not to fall for phish. *ACM Transactions on Internet Technology 10*, 2 (May 2010), 1–31.

[20] LENNON, E. B. *IT security metrics*. August 2003.

[21] ORME, J. G., AND COMBS-ORME, T. *Multiple Regression with Discrete Dependent Variables*. 2009.

[22] PAXSON, V. Strategies for sound internet measurement. In *IMC '04: Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement* (New York, NY, USA, 2004), pp. 263–271.

[23] RITCHIE, F. Secure access to confidential microdata: Four years of the virtual microdata laboratory. *Economic and Labour Market Review 2*, 5 (May 2008), 29–34.

[24] ROSSOW, C., DIETRICH, C. J., KREIBICH, C., GRIER, C., PAXSON, V., POHLMANN, N., BOS, H., AND VAN STEEN, M. On the soundness of silence: Assessments of malware execution studies. In *Proceedings of the 33rd IEEE Symposium on Security and Privacy* (2012). to appear.

[25] SECURITY, U. D. O. H. *A Roadmap for Cybersecurity Research*. 2009.

[26] SHARMA, A., KALBARCZYK, Z., IYER, R., AND BARLOW, J. 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, 2010), pp. 144–151.

[27] SHENG, S., HOLBROOK, M., KUMARAGURU, P., CRANOR, L. F., AND DOWNS, J. Who falls for phish?: A demographic analysis of phishing susceptibility and effectiveness of interventions. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems, CHI '10* (New York, NY, USA, 2010), pp. 373–382.

[28] SHENG, S., MAGNIEN, B., KUMARAGURU, P., ACQUISTI, A., CRANOR, L. F., HONG, J., AND NUNGE, E. Anti-phishing Phil: The design and evaluation of a game that teaches people not to fall for phish. In *SOUPS '07: Proceedings of the 3rd Symposium on Usable Privacy and Security* (New York, NY, USA, 2007), ACM, pp. 88–99.

[29] SOMMER, R., AND PAXSON, V. Outside the closed world: On using machine learning for network intrusion detection. In *Proceedings of the 31st IEEE Symposium on Security and Privacy* (2010), pp. 305–316.

[30] STONE-GROSS, B., COVA, M., CAVALLARO, L., GILBERT, B., SZYDLOWSKI, M., KEMMERER, R., KRUEGEL, C., AND VIGNA, G. Your botnet is my botnet: Analysis of a botnet takeover. In *Proceedings of the 16th ACM Conference on Computer and Communications Security* (New York, NY, USA, 2009), pp. 635–647.

[31] WEIR, M., AGGARWAL, S., COLLINS, M., AND STERN, H. Testing metrics for password creation policies by attacking large sets of revealed passwords. In *Proceedings of the 17th ACM Conference on Computer and Communications Security* (New York, NY, USA, 2010), pp. 162–175.

[32] WIMBERLY, H., AND LIEBROCK, L. M. Using fingerprint authentication to reduce system security: An empirical study. *In Proc. of the IEEE Symposium on Security and Privacy 0* (2011), 32–46.

[33] WINTERS, P. R. Forecasting sales by exponentially weighted moving averages. *Management Science 6*, 3 (1960), 324–342.

[34] YOUNG, J. R. *Academic publisher steps up efforts to stop piracy of its online products. The Chronicle of Higher Education.* `http://chronicle.com/article/Academic-Publisher-Steps-Up/128031`. June 26, 2011.

[35] ZHANG, J., BERTHIER, R., RHEE, W., BAILEY, M., PAL, P., JAHANIAN, F., AND SANDERS, W. Safeguarding academic accounts and resources with the university credential abuse auditing system. In *Proc. of the IEEE/IFIP International Conference on Dependable Systems and Networks (DSN 2012)* (Boston, MA, June 2012).