Annales UMCS Informatica AI 2 (2004) 413-424

Annales UMCS
Informatica
Lublin-Polonia
Sectio AI

http://www.annales.umcs.lublin.pl/

Developing an empirical study of how qualified subjects might be selected for IT system security penetration testing

L.M. Guard[a], M.D. Crossland[a], Marcin Paprzycki[bc*], J.P. Thomas[b]

[a]*Management Science and Information Systems Department*
[b]*Computer Science Department, Oklahoma State Uniwersity, Tulsa, OK 74106, USA*
[c]*Department of Mathematics and Computer Science, Adam Mickiewicz University, Poznań, Poland*

Abstract

This paper describes a planned program of investigation designed to determine what characteristics are signficant in predicting performance of students used as subjects in IT system penetration testing testbeds. In large part the experimental design replicates an earlier study by Jonsson et al., and extends that study to include factors describing the attacking subjects. In this way the proposed study is expected to be able to verify and further their work by collecting data on a larger population of subjects. Among others we expect to verify their hypothesis that to the stationary nature of the breaking-in process and the intrusion process during the standard attack phase is characterized by exponential distribution. Finally, the proposed study will be also usable for the purpose of evaluation of intrusion detection systems.

1. Introduction

Following the September 11, 2001 attacks a need has been recognized for establishment of a large number of testbeds for conducting security related information technology (IT) evaluations. Such efforts are funded by the US government [1,2] as well as business (e.g. Microsoft [3]) and the funding can be expected to further increase. In this context one of the important issues is ensuring an adequate supply of penetration testers to evaluate the security of IT systems. While some researchers (Puketza et al., [4]), propose an approach in which security research is based on the computer simulation of attackers, others suggest that such a research should utilize humans. Furthermore, it is argued that instead of expert hackers, "normal users" are better suited for the IT system security modeling [5]. In this context four observations can be made. First, universities are capable of providing a large number of students to work as

* Corresponding author: *e-mail address*: marcin@cs.okstate.edu, work at Adam Mickiewicz University was sponsored by a scholarship from the Fulbright Commission.

testbed experimenters. Second, students work very well as representatives of "typical users" [5]. Third, creation of university-located testbeds matches the substantial increase in US governmental spending intended to increase the number of graduate students focused on information assurance [6]. Fourth, many university computer/network labs could be easily turned into a controlled environment and utilized as a testbed location. Furthermore, such labs have the necessary facilities, equipment and software for setting up a wide variety network configurations and target systems. Finally, university labs can be easily and cheaply separated from the Internet to assure safe conditions for penetration testing (as well as other IT security related research).

In a university based IT security testbed research can be conducted to pursue two important goals: (a) to mimic realistic security penetration situation and thus to evaluate the security of a given IT system and (b) to utilize the same behavior to evaluate the intrusion detection systems (IDSs) [7-10].

Goal (a) should be viewed from two perspectives. A recent article by Jonsson et al. [5] set forth a methodology to construct a quantitative model of the behavior of attackers against an IT system. Since their experimental data was collected on a very limited scale (only 12 groups consisting of 2 students each), an ability to further verify their findings on a larger group of subjects would have important consequences to developing quantitative measures of IT system reliability (similar to these existing in other engineering disciplines). This is especially the case since the authors of [5] themselves posed a number of questions and open hypothesis that according to them merit further explorations. On the other hand, Jonsson et al. [5] pointed out that some of the student attackers have a very long learning curve prior to accomplishing any security breach, while other students never achieve a breach. It can be stipulated that students that are unable to accomplish a security breach even after receiving training in penetration testing have no business working in most testbed experiments as attackers, and they take up valuable space in the training program. This problem could be avoided if there existed a set of criterion that could be utilized to screen students seeking admission into a testbed program that would help ensure that most of admitted students can become useful for the research.

Goal (b) is equally important. Currently there are two major approaches used by IDS to detect intrusive behavior: anomaly detection [7,8] and pattern-matching detection, a/k/a misuse detection, a/k/a model-based detection [10]. While some IDS utilize both types of approaches and others utilize rule-based expert systems to perform or assist in the analysis, most IDSs can be easily classified into one or both of these two categories [10]. Each type of IDS has its own strengths and weakness and is technically suited to identify only a specific subset of known security violations [8]. This reflects the need for and the

importance of all types of IDS research; while the research involving human attackers working in a controlled environment becomes particularly valuable.

This paper describes a planned study that is to address all of the above issues. First, it will in some respects replicate the study reported in [5] and lead to further verification of results obtained there as well as deep probing of some open questions posed by its authors. These questions were unanswered due to the limited supply of experimental data. Second, it will try to discover important and easily measurable factors that influence future success of a student as security expert that can be utilized in the future to pre-screen candidates for (rather expensive) security training. Third, it will initiate collection of data related to effectiveness of various intrusion detection systems. Furthermore, we plan to achieve all these three goals by training one group of students and analyze their behavior as "hackers".

2. Previous research and theory

In their paper [5], Jonsson et al. presented the results of an experiment conducted with the goal of expressing security quantitatively. They studied system's ability to avoid influence created by a hacker, and obtained measures of the preventive characteristics of a system. To achieve this goal and to formulate a theory of attacker behavior, which constitutes a basis for a way of modeling the intrusion process, they collected time-related data reflecting the amount of time the attackers utilized during each breach of security they accomplished. Their main results were: (1) that the intrusion process could be split into three phases: a learning phase, standard attack phase, and innovative attack phase and (2) that the intrusion process during the standard attack phase could be described by an exponential distribution curve.

This latter result indicated that traditional methods for reliability modeling used in many engineering disciplines, e.g., Markov models, might be applicable in security modeling. This led, for instance, to application of a Markov model to compute security measures (Ortalo et al. [11]). The experiment reported in [11] confirmed the expected behavior of the Markov model with respect to the attacker model assumptions.

2.1. Purposes of the experiment

The primary initial purpose of the proposed experiment is to attempt to solve a real world problem of designing a predictive instrument for the selection of students for penetration testing training & testbed work. It is believed that there exist certain identifiable individual background experiences, attributes, which could be utilized to establish those types of prerequisites.

In an effort to address the problem, this experiment will attempt to determine what level, if any, of the below identified attributes have significant positive

interactions with a student's ability to learn the basics of penetration testing in a relatively short time frame. Outside of intelligence, it is believe that IT related classes and experiences have the strongest correlations with a student demonstrating penetration testing mastery. However, since this portion of the study is exploratory, we intend to cast a wide net and will look at the broadest possible range of the students' backgrounds. We will gather complete transcripts from all students and obtaining detailed background information which will all be analyzed. In addition to the specific IT related backgrounds identified below, we intend to look, among others, at such issues as total credit hours and GPA accumulated in humanities, social sciences, and mathematics. We will also look at where they went to high school, their age, gender, ethnicity, learning styles and personality types.

A second purpose of the experiment is an attempt to further investigate the Jonsson et al. theory that the times to breach of security in an IT system caused by the actions of attackers in the standard attack phase are exponentially distributed [5]. Part of this investigation involves looking for possible alternative explanations suggested by Jonsson et al. on the important issue of whether the security breach process is stationary, i.e. the time between security breach n and (n + 1) is independent of the time between breach (n + i) and (n + i + 1) for all i's. This is the third necessary precondition for the test of whether the times to breach are exponentially distributed. While Jonsson et al. tested for the validity of this assumption and found no reason to reject it, they also noted two alternative explanations which may have been present but the effects of which were simply too small to be detected from their sample data. In the experiment reported in [5] there were only 24 students involved and they were divided into 12 groups of two students. Since only group data was analyzed, the experiment was based on a sample that was simply too small for any substantial generalization. By utilizing more student attackers in the planned experiment, by having the students work independently, and by designing this experiment and the data collection processes in such a way as to allow direct comparison of the results of this experiment with the results from the Jonsson et al. experiment, it is hoped that we will eventually collect enough data (that is, to obtain a large enough N), to eventually detect the critical assumption killing effects if they are present at all.

Finally, we also intend to evaluate multiple IDSs during this study. The final selection of which IDSs to evaluate has not yet been made. All of the IDSs will be deployed without the students knowing anything about them and they will be installed on computers with non routable IP addresses to help prevent the students from finding them during the study. IDSs will be configured to log all network activity during the study and to send notifications to the lab network administrator ever time an event occurs that potentially represents a threat to the

target system. This will provide means to verify what the students do during the study as well as basis for comparing the performance of the IDSs.

2.2. Research hypotheses

As indicated above, we plan to cast a wide net at considered factors that can influence the performance of student penetration testers. However, there are some basic hypothesis as to which factors are likely to play the crucial role. In the following hypotheses, the Total Grade Points Earned (hereinafter "TGPE") is derived from a student's academic background, and is calculated as the product of number of credits hours earned times the points of the grade earned in certain types of courses. Further explanation can be found in section 3.5.

- **H1**: Students with higher TGPE in programming languages will penetrate targeted systems in less time than those with lower TGBE in programming languages, including those with no instruction.
- **H2**: Students with higher TGPE in networking or telecommunications will penetrate targeted systems in less time than those with lower TGBE networking or telecommunications, including those with no instruction.
- **H3**: Students having experience with more operating systems administration will penetrate targeted systems in less time than those with less experience in operating systems administration, including those with no experience.
- **H4**: Students with higher TGPE in information security classes will penetrate targeted systems in less time than those with lower TGBE in information security, including those with no instruction.

3. Methods

3.1. Summary of the procedures, methodologies and processes comprising the study

1. Volunteer undergraduate and graduate student subjects will be solicited from a large Midwestern United States university as well as from a computer forensics program at a local community college in an effort to obtain the widest possible backgrounds in the students. Those selected for the study will be at least in their second year of undergraduate studies with some IT related coursework or experience.
2. Subjects who are selected will provide certified copies of all college transcripts, all college entrance exam scores, complete a Subject Background Experience Report and evidence of any other IT related training and background. Additionally, all students will complete short instruments designed to assess personality types and learning styles. Final selection of these instruments has not been made but they will be chosen from previously validated measures.

3. Subjects will first be divided into groups based on the quantity and quality of their IT backgrounds. As many groups will be utilized as appears reasonable after the data is analyzed. Then the students will be randomly assigned, in approximately equal numbers from those groups, into a treatment and control group.
4. Subjects in the treatment group will be provided with an online two-day introductory penetration testing training seminar based on the substantive subject matter in [12]. Subjects in the control group will receive an online introduction to their objectives and tasks they are to perform, plus only basic instructions in how to operate the software. However, none of the substantive material from [12], like an overview of network protocols, will be taught in the control group on-line seminar. This is expected to allow us to see if there are differences based on what the students know before they start the study and how well the students can learn the substantive material taught during the study.
5. Subjects will then be provided with the testbed facilities including computers for them to launch their attack from and a variety of target IT systems for them to attack. Students will also be provided a copy of [11] to utilize in the lab and copies of all software discussed in [12]. During the study, each student will be required to work independently for a total of 40 hours.
6. An observer will record the time each student starts working in the testbed and the time of, and nature of, each penetration created by each student. Subjects will also complete Attacker Activity Reports documenting what they do and the time involved.
7. Utilize two IDSs during the study to record all the students' network-based activity for analysis in order to compare the accuracy of each.

3.2. Selection and measure of the dependent variable

Jonsson et al. looked at many options for what they were going to utilize as the primary variable in their experiment. According to Jonsson, et al.:

> The criteria for selecting a good variable ...: first, it should be a variable that captures our intuitive notion of the breach process, and it should be suitable and make sense in a measure such as – mean time to breach, MTTB. It must also be a variable that gives usable results in the modeling work [5].

Of the options Jonsson et al. discussed, group working time (t_{gw}) was selected as their primary variable for modeling the intrusion process. Group working time, t_{gw}, was equal to all preparation time, t_p, plus all attack time, t_a, utilized by each group member both when working alone and when working as a group. Therefore, $t_{gw} = t_A + t_B + t_{A+B}$.

In this planned study the DV will be each student's average of the Jonsson et al. time variable. While Jonsson et al. provide a reasonable explanation as to why they allowed students to work in groups of two, it was decided that the students in this study will be required to work individually so as to avoid any potential contamination of the effects of a student's individual attributes by virtue of working with a partner with different attributes. In this study the DV will be the individual's average working time, t_w, and the average amount of time a student spends accomplishing each security breach. The student's working time, t_w, for each breach will be all preparation time, t_p, plus all attack time, t_a. More succinctly, $t_w = t_p + t_a$.

3.3. Selection of student attackers

It could be argued that some types of penetration testing security assessments performed in testbeds should be performed by professionals or at least individuals with skill and knowledge levels of professionals. However, utilizing students as attackers in the evaluation of an IDS and many other situations make good sense. It is much less expensive to utilize students as attackers in a testbed instead of professionals. Additionally, in many situations scheduling problems greatly complicates the use of professionals. Jonsson et al. decided to use students in their experiment and they explained why they did not want to use professionals:

> We were aiming for attackers that could be considered to be the "normal" users of the system, i.e., users without any special knowledge of security issues. It is important to note that we did not want professional crackers who already knew about most weaknesses in the system. Professional crackers would give us information only about where and now our particular system needed to be improved. Such experiments or investigations have indeed been performed (citations omitted). However, we had to use "normal" users attacking the system, first in order to study the intrusion process in detail to be able to present a model for the intrusion process, and second, to see how well the system could withstand attacks from its "normal" users, and hopefully to present at least a crude measure of, for example, mean time to breach. Such a measure would be extremely useful for both the system owner and for its users when deciding what information could be stored on the system [5].

It should be stressed that while the reasons of Jonsson et al. were compelling for their experiment, the reasons they decided to use students are not the same as in this study. However, their reasons make it clear that it has been recognized that depending on the purpose of the study, students can actually be preferred over professionals for use as attackers in a testbed.

3.4. Selection of treatment – penetration testing training program

In a predictive or correlation study, such as this one, there is no need for multiple treatments. However, determining what the training should be comprised of and how the material should be presented was a challenge. Based on conversations with students it is believed that it will be the training that will be real inducement for the students to participate in the study. This is because this type of training is very expensive in the commercial world; US$2,600 for a two to four day course on penetration test is typical [1].

Due to concerns for low student participation rates from students knowing there was a 50% chance they would not get the training, and high differential mortality in the control group from students knowing for certain they will not get the training, the idea of using a control group that never receives the training was rejected. Furthermore, the idea of two fundamentally different training programs being used was rejected due to concerns relating to the conclusions that could be drawn from the analysis of the data. It was finally decided that there would only be one training program and a control group, but the control group will receive the training following the study. Additionally, the control group will be provided with the opportunity to return to the testbed to practice any techniques or try any software they did not successfully use during the study. In this way, we can honestly tell the students that they may be better off being assigned to the control group if what they want is to learn as much as possible.

3.5. Selection of independent variables

The following are the subject attributes that will be measured as the IT related independent variables. Since it is anticipated that all subjects will come from a student population, the variables selected deal with the subjects' number of credit hours of study and grades obtained in a number of academic subjects. The first three of the following variables will be operationalized as the product of number of credits hours earned times the points of the grade earned in that course. The product is referred to as Total Grade Points Earned (hereinafter "TGPE"):

1. Programming language instruction, including BASIC, Visual BASIC and languages used for Common Gateway Interfaces;
2. Information security instruction including computer forensics; and
3. Networking and telecommunications management and/or protocol instruction.

The following variable will be operationalized as simply a count of the number of operating systems with which the subject has experience:

1. Operating systems administration experience – students will only receive credit for each operating system they can install in a network environment and configure for a user.

It is realized that students would not necessarily acquire all their potential attributes through classes. Therefore, the measure of attributes in students will not be limited to class work as long as there is means available to verify the existence of the attribute, e.g., a Microsoft certification, or work experience. A student's experience will be assessed by the researcher, who will decide how many "TGPE equivalency" to provide for the experience.

There will also be exploratory analysis conducted on a wide range of non IT related attributes including but not limited to total GPA, GPA in social sciences, GPA in humanities, GPA in mathematics, personality types and learning styles.

3.6. Selection and measure of control variable

It is believed that students' intelligence will have a positive correlation to their ability to demonstrate a mastery of penetration testing, but it would be inappropriate to utilize an intelligence related perquisites for admission into a university class. Therefore, intelligence will be utilized as a control variable in the study. Intelligence will be operationalized by using a proxy, the students' standardized ranking on their SAT or ACT exams, like was done in [13]. The correlation between college entrance exam scores and IQ scores has been provided by [14].

Our initial investigation will analyze each exam type separately across the study. However, the main objective is to draw comparisons across the entire set of subjects, and it is anticipated that either type of exam score can reasonably be used as a proxy for intelligence. Because the SAT score scale is different than the ACT score scale, all SAT composite scores therefore will be converted to the correlated ACT composite scores for standardization of the comparisons across the study. The College Board provides tables for comparing SAT and ACT scores [15].

4. Critique of the experiment

There are a number of issues that may be raised with regard to our proposed study and we are well aware of them. Let us now discuss the most important critical issues.

1. The attribute interval scales selected for this experiment lack true isomorphism. This is because for all background experiences gained outside of a classroom the only control for standardizing the scale is by virtue of having one person, make all final decisions regarding the number of how many "TGPE equivalency" to provide for the experience. However, we have decided that establishing the TGPE may be one of the few reliable ways of measuring level of "professionalism". Since one person will evaluate all students we believe that we will be able to ensure consistency in establishing TGPE.

2. The measure of the attributes do not capture the true complexity of the constructs we are interested in. This is because there is no reason to believe that all students that earned the same grade for what appear to be equivalent classes possess equivalent background experiences as of the time of the experiment. To the contrary, it is believed that both how long ago a class was taken and who the instructor was would make a real differences. Unfortunately, we do not see any practical way of tackling this problem.
3. While the time variable in the DV has been previously utilized for a very similar purpose, it cannot be considered ideal. Security breaches are too complex for a student's ability to create them to be totally and repeatably captured by a simple time variable. The measure of the DV would be improved if the different degrees of difficulty in causing specific breaches could be ranked and weighted. Then the weights could be utilized in the measure of the DV in conjunction with the time variable. Thus, while a better measure of the DV may be possible, none exists at this time, and it would certainly take time to create one. In addition, since we would like to be able to further the research initiated by Jonsson, we decided that our selection of the DV is the best possible one at this stage. However, since all of the experimental data will be collected by the IDS's we will be able to return to it in later stages of our research and pursue the proposals made here.
4. A potential problem exists in the selection of the treatments. The control group is going to be provided the same book and software as the group that goes through the training. The fact that both sets of students will be utilizing the same resource material and tools during their attacks on the target system could create some covariation between the two groups DVs. However, it was decided that this potential problem was more than offset by other benefits, including: by choosing to leave this risk in place, we have all conditions for both groups identical except the focus of the training; it is considered a strong possibility that many students, especially those with low values for the selected attributes, would have little chance of creating many, if any security breach, regardless of how insecure the target system was, if they did not have a good and easily available resource to provide them with ideas on how to conduct an attack. Thus, while this potential problem still exists, it is believed that there will be a strong enough difference in the DVs of the two treatment groups that the best possible decisions for the treatments in this experiment have been made.

A potential finding pursuant to item #4 above may be that the additional, concentrated training provided to the experimental treatment group will not provide any significant contribution to student attackers' success in accomplishing a penetration. If this is found to be the case, then perhaps it may

be concluded that spending the additional time and effort to train student attackers is not worthwhile, or that some other form of training could or should be substituted for future studies. The latter conclusion might be warranted if no other independent variables are found to significantly contribute to penetration success.

5. Implications of findings

If our research hypotheses are supported we will have at least found a reasonably good solution to the pressing real-world problem of how to establish a set of objectively verifiable criteria that can be utilized as prerequisites for student admission into a testbed penetration testing program which assures that most of students admitted into the program should be able to master the basics of penetration testing in a relatively short time.

The results of the study are expected to be rich enough that we will not be limited to providing a single solution to the problem, but instead a set of potential solutions which can be customized for a given testbed project. Additionally, we should be able to establish at least one set of more stringent prerequisites which would be appropriate when higher levels of penetration testing proficiency is considered essential for the attackers. These, in turn, can possibly suggest modifications to the programs of study offered by various IT security assurance programs – by identifying which factors are the most important we will be able to propose how to make sure that precisely these factors are the central part of the curriculum.

6. Concluding remarks

The results of this proposed study are needed right now. Cybercrime and threats of cyberterrorism leave us with no option but to consider the field of information assurance vitally important to the future of the IT infrastructure around the world. One essential aspect of information assurance is the testing of the performance of IDS as well as other IT hardware and software. When this testing is performed by a third-party independent from the manufacturer, it is normally performed in a testbed. This is the reason the federal government perceives a need for a significant increase in the number of testbeds in US.

If significant numbers of additionally testbeds are to be created and operated at university it is essential to staff them with model-attackers in some economical way and training students to work as attackers is an attractive option. It has already been demonstrated that some form of screening for potential student workers in testbeds is necessary to avoid the inefficiencies of having student attackers working in a testbed that are incapable of creating a security breach in a target system. This proposed study is believed to be a good step

toward the establishment objective criteria that can be utilized to screen students for admission into testbed training programs.

Studies of this type may contribute toward a goal of making security a structured and reliable discipline, similar to those in engineering, particularly software engineering. The intended types of training and testbed work may be able to contribute to professional certification needs in this important field.

References

[1] Foundstone, Special Edition Classes, 2-4 day Ultimate Hacking: Hands On courses cost $2,600, http://www.foundstone.com/services/special_edition_classes.html, last accessed April 22, (2002).

[2] Furlani, Cita, Director of the National Coordination Office for Information Technology Research and Development, "Federally-Funded information Technologies," http://www.itrd.gov/about/presentations/furlani-thunderbird-16jan01/index.html, last accessed April 22, (2002).

[3] Evers Joris, *Microsoft Asks Colleges to Teach Hacking, Students will learn how to hack into software and fix its bugs,* http://www.pcworld.com/news/article/0,aid,109935,tk,dn032103X,00.asp, last accessed October 18, (2003).

[4] Puketza, Nicholas J., Zhang, Kui, Chung, Mandy, Mukherjee B., *A Method for Testing Intrusion Detection Systems*, IEEE Transactions on Software Engineering, 22(10) (1996) 719.

[5] Jonsson, Erland and Olovsson, T., *A Quantitative Model of the Security Intrusion Process Based on Attacker Behavior*, IEEE Transactions on Software Engineering, 23(4) (1997) 235.

[6] Marsan, Carolyn, *Congress: Tighten IT Security*, NetworkWorld, 19(16) (2002) 1.

[7] Hill, John M.D., Carver, Jr., Curtis A., Humphries, Jeffrey W., Pooch, Udo W., *Using an Isolated Network laboratory to Teach Advanced Networks and Security*, ACM SIGCSE Bulletin, Proceedings of the thirty-second SIGCSE technical symposium on Computer Science Education, 33(1) (2001) 36.

[8] Ilgun, Koral and Kemmerer, Richard A., *State Transition Analysis: A Rule-Based Intrusion Detection Approach*, IEEE Transactions on Software Engineering, 21(3) (1995) 181.

[9] Northcutt, Stephen, Network Intrusion Detection: An Analyst's Handbook, New Riders Publishing, Indianapolis, IN, (1999).

[10] Tipton, Harold F., Krause M., Editors, Information Security Management Handbook, 4th Ed., CRC Press LLC, Boca Raton, FL, (2001).

[11] Ortalo, Rodolphe, Deswarte, Yves and Kaaniche, Mohamed, *Experimenting with Quantitative Evaluation Tools for Monitoring Operational Security*, IEEE Transactions on Software Engineering, 25(5) (1999) 633.

[12] McClure, Stuart, Scambray, Joel, Kurtz, George, Hacking Exposed: Network Security Secrets & Solutions, 4th Ed., McGraw-Hill Osborne Media, Berkeley, CA, (2003).

[13] Applin, Anne Gates, *Second Language Acquisition and CS1: Is * == ** ?*, ACM SIGCSE Bulletin, Proceedings of the thirty-second SIGCSE technical symposium on Computer Science Education, 33(1) (2001).

[14] Benet W.E., *IQ Resources*, http://webenet.com/iq.htm, last accessed Nov. 9, (2003).

[15] College Board 2000 The College Board, "2000 SAT I-ACT Score Comparisons", http://www.collegeboard.com/sat/cbsenior/html/stat00f.html, last accessed Nov. 9, 2003.