Communications of the Association for Information Systems

Volume 14 Article 32

December 2004

Empirical Test of a Hacking Model: An Exploratory Study

Al Bento
University of Baltimore, abento@ubalt.edu

Regina Bento
University of Baltimore, rbento@ubalt.edu

Recommended Citation
Bento, Al and Bento, Regina (2004) "Empirical Test of a Hacking Model: An Exploratory Study," *Communications of the Association for Information Systems*: Vol. 14 , Article 32.
DOI: 10.17705/1CAIS.01432
Available at: https://aisel.aisnet.org/cais/vol14/iss1/32

Communications of the Association for Information Systems

EMPIRICAL TEST OF A HACKING MODEL: AN EXPLORATORY STUDY

Al Bento
Regina Bento
University of Baltimore
abento@ubalt.edu

ABSTRACT

This exploratory study is an empirical test of a model of the activities involved in hacking attacks, and the conditions associated with the increase in these activities. In a methodological innovation, the variables in the model were measured using non-reactive, secondary data obtained from sixty months of official statistical data, from 1999 through 2003. These variables were analyzed using stepwise regression. The results obtained support several of the model predictions.

1. Increased Broadband access by home and small business users is associated with an increase in Reconnaissance activities by hackers looking for vulnerable systems.
2. Increased Reconnaissance is associated with an increase in hacking attempts to obtain initial access through the use of Malicious Code.
3. The increase in User Compromise is associated with the increase in Root Compromise, reflecting hackers' efforts towards escalation of privilege.
4. The negative relationship between Root Compromise and Denial of Service supports the prediction that hacker frustration at failing to gain control of a resource may be one of the factors contributing to Denial of Service attacks.
5. Environmental variables (Broadband and Number of Hosts) are positively related to one another.

These results suggest potentially significant implications for research and practice.

Keywords: hacking, empirical test, security, network, Internet

I. INTRODUCTION

Empirical research on hacking is still in its early stages, despite widespread public concern with network and Internet security. Most empirical work on computer security predates the major growth of the World Wide Web in the mid-1990's [Bookholdt, 1989; Loch et al., 1992; Straub and Nance, 1990], and thus fails to take into consideration the new dimensions that the Web added to computing security. This exploratory study proposes a model of the activities involved in hacking attacks, including the impact of environmental variables such as the growth in the number of

Internet hosts and the increase in broadband access. Sixty months of official statistical data (from 1999 through 2003) were used in the preliminary test of the model.

Networks and the Internet are decades old, but it was the advent of the World Wide Web in the 1990's that made them pervasive in businesses and homes. The significant growth of the Web led to the realization that the Internet poses an ever-increasing security threat [Straub and Welke, 1998]. In the past, security professionals believed that most attacks on computers and networks came from inside the organization, but the growth of the Web turned outsiders into the biggest threat [Pfeegler and Pfeegler, 2002].

The CERT Coordination Center[1] registered a significant growth in computer security incidents: the number of incidents reported increased from 9859 in 1999 to 137,529 in 2003 [CERT/CC, 2004]. Trends include a continual increase in the speed and sophistication of attack tools, faster discovery of vulnerabilities, increasing permeability of firewalls, increasingly asymmetric threat, and increasing threat from infrastructure attacks [CERT/CC, 2002]. A 2003 survey by the Computer Security Institute and the Federal Bureau of Investigations (CSI/FBI) found that 75% of the 530 organizations in the CSI/FBI sample detected computer security breaches leading to financial losses [Richardson, 2003]. In spite of the difficulties involved in measuring the costs of cybercrime, those costs are estimated to be substantial [Garg et al., 2003] and may grow at yearly rates of about 200% [Lukasik, 2000]. Given the tendency for cybercrime to be underreported, the magnitude of the importance of security breaches may be even greater than the bleak scenarios and trends already identified [Bagchi and Udo, 2003].

Security breaches receive intense media attention (e.g., the hoopla surrounding attacks such as *Melissa* and *Nimda*). The trade literature offers substantial information about hacking methods, techniques, tools and countermeasures [McClure et al., 2001]. The academic literature on security breaches is mostly limited, however, to case studies about specific incidents or organizations [Straub and Welke, 1998] and analytical studies based on surveys of experienced attacks [Bagchi and Udo, 2003]. Few previous studies used actual data about security incidents reported to CERT/CC [Howard, 1997].

In the sections that follow, we present a brief overview of a hacking framework (Section II), describe the model we developed to understand the factors that contribute to the various activities involved in hacking attacks (Section III), and the research questions and hypotheses inspired by the model (Section IV). In Section V we discuss our data collection strategy, which used surrogate measures based on official, publicly available statistics from the Federal Computer Incident Response Center (FCIRC), the Federal Communications Commission (FCC), and Netcraft. Section VI presents an analysis of the results of the empirical test of our model. We conclude by discussing, in Section VII, the implications of this exploratory study for research and practice.

II. HACKING FRAMEWORK

The hacking framework presented here includes elements widely used in the development of practical tools for prevention and defense against hacking attacks [Bento, 2003; Howard, 1997; Howard and Longstaff, 1998; McClure et al., 2001; Panko, 2003; Pfleeger and Pfleeger, 2002]. The framework identifies four steps in hacking attacks: (1) information gathering; (2) initial access; (3) privilege escalation; and (4) covering tracks and creating back doors.

[1] Established in 1988, the CERT® Coordination Center (CERT/CC) is a center of Internet security expertise, located at the Software Engineering Institute, a federally funded research and development center operated by Carnegie Mellon University.(Source:CERT Coordination Center http://www.cert.org/ (Current November 29, 2004) The Center tracks computer security incidents.

STEP 1. INFORMATION GATHERING FOR TARGET ACQUISITION

The information gathering activities in Step 1 can be compared to a burglar "casing the establishment" [McClure et al., 2001, p. 1]. They involve three main types of activities: footprinting, scanning, and enumeration.

Footprinting

The goal of footprinting is to gather as much information as possible about all aspects of the target and its security, just as a bank robber would try to find out about such things as armored car routes and delivery times, video cameras, number of tellers, escape routes,. [McClure et al., 2001, p. 2]. Using a variety of techniques, the attacker uncovers information about the target's environments such as IP addresses reachable through the Internet, TCP and UDP services, system architecture, access control mechanisms, remote access, and extranet.

Scanning

Scanning allows the attacker to focus on those systems in the target that are "alive" and actually reachable through the Internet, just as a burglar would try to find out which doors or windows are most vulnerable [McClure et al., 2001., p. 30]. Scanning involves techniques such as "ping sweep" (to find which IP addresses have active hosts), TCP/UDP port scanning (to find out which ports have active server programs running), and operating systems detection.

Enumeration

Enumeration refers to an intrusive probe to identify valid user accounts, network resources and shares not adequately protected, applications and versions, and other details. Enumeration activities can potentially be logged and detected, given their intrusive nature, with active connections to systems and directed queries.

STEP 2. INITIAL ACCESS

Step 2 includes attempts to gain access to the target's system, and to compromise it as much as possible, after obtaining user-level privileges.

Attempts to gain access often involve malicious code, such as viruses that infect files in a single system, worms that spread infections across different systems (such as *BubbleBoy*, and *I Love You*), and blended worms or snakes, which can carry viruses and "Trojan horses" (*Code Red, Code Red II, Sircam, Nimda*). A "Trojan horse" (or "Trojan," for short) is a program that pretends to be legitimate software, such as a game, but "performs unintended (and often unauthorized) actions, or installs malicious or damaging software behind the scenes when launched." [McClure et al., 2001, p. 578]. Hackers may also attempt to gain access to a user's system by using techniques such as brute force password guessing and buffer overflows [Panko, 2003, p. 315; McClure et al., 2001, p. 161]. Still another approach is to gain physical access to the user's system (e.g., when computers are left unattended in the workplace), but this method tends to be a less frequent form of initial access.

Once hackers succeed in "opening the door" to the user's computer (through malicious code, brute force. or physical access), they then proceed to breach its security and compromise its confidentiality, integrity and availability [Pfeegler and Pfeegler, 2002].

STEP 3. ROOT COMPROMISE

In step 3 the attacker tries to gain complete control over the system by acquiring privileges above the simple user-level. Control can be achieved directly or indirectly, starting with User Compromise and then achieving Root Compromise through escalation of privilege. The hacker tries to acquire administrator or root privileges (through techniques such as password cracking and Trojans), and to consolidate power by obtaining other accounts, and accessing other resources (hosts or networks). The hacker is now in a position to wreak havoc in the system by means such as reading or altering sensitive information, changing or deleting key files, wiping out the hard drive, and using the compromised target to launch attacks against other targets. [Panko, 2003].

STEP 4. COVERING TRACKS AND CREATING BACK DOORS

In Step 4 the hacker takes advantage of the administrator-level control of the target that was acquired in Step 3, to try and avoid detection by the system's own administrators. Techniques include deleting or modifying logs, hiding tools, and disguising Trojans.

The hacker may also set backdoors, to ensure that access can easily be regained later, even if the password is changed. Techniques include creating rogue user accounts, scheduling batch jobs, infecting startup files, planting remote control services, installing monitoring mechanisms and replacing applications with Trojans.

STEP SEQUENCING, ALTERNATIVE PATHS AND DENIAL- OF- SERVICE

The framework implies a logical sequence of steps: gathering information, then breaking in to gain user-level access, then using this level of access to gain higher level privileges, and finally covering the tracks and leaving backdoors open for return intrusions. It is important to note, however, that some steps may be skipped (e.g., gaining access without having bothered to gather information; or obtaining administrator privileges already in the initial access) or repeated (e.g., going back for more elaborate enumeration after gaining administrator privileges).

Not all attacks succeed. When hackers are unable to achieve control of the target, they often express their frustration by launching Denial-of-Service (DoS) attacks (such as *Smurf, Fraggle* and *Syn*) that disrupt services or make them inaccessible to legitimate users, networks, and systems. Techniques for DoS attacks may involve bandwidth consumption, resource starvation, taking advantage of programming flaws, and launching routing and DNS attacks. An even more vicious form of attack is Distributed Denial-of-Service (DdoS), where handler programs and zombies or slaves are planted in several other compromised clients or servers, which are then used to attack the target [McClure et al., 2001, p. 504]. This form of attack succeeded in temporarily paralyzing big multiuser targets such as Yahoo, eBay, CNN.com, E*Trade, ZDNeT and others, causing severe financial losss.

III. THE MODEL

Figure 1 presents the model we developed to test the hacking framework. It provides a basis for exploring the variables inspired by the steps in the framework (reconnaissance, malicious code, user compromise, root compromise, denial of service). It also adds two other variables that might have an impact on the growth in security breaches: number of hosts in the Internet, and broadband access to the Internet by home and small business users.

RECONNAISSANCE

This variable corresponds to Step 1 in the hacking framework (information gathering for target acquisition, through footprinting, scanning and enumeration). The objective of reconnaissance activities is to identify potential victims for future hacking attacks, by obtaining information such as number and type of computers, operating systems, servers, applications, and resources such as shared files, and databases. If the hackers find enough interesting resources in a given site or organization, they are then more likely to attempt initial access.

MALICIOUS CODE

Malicious Code is the tool of choice for hackers trying to gain initial access to user and administrator accounts (the beginning of Step 2 in the hacking framework). As discussed in Section II, Malicious Code attacks include worms, viruses, and similar computer code, and

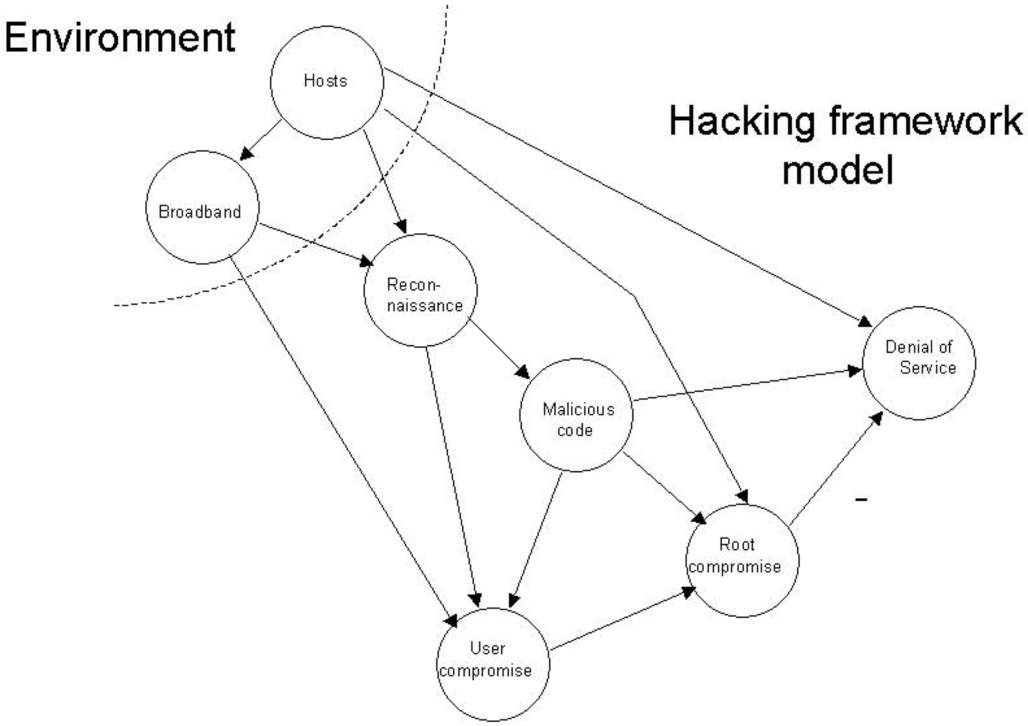

Figure 1. Model for Empirical Test of Hacking Framework

typically deliver Trojan horse payloads to a target's computer. Malicious Code attacks may exploit software vulnerabilities in popular operating systems, server software and application software, or rely on visit to a web site which delivers the malicious code directly. These attacks may also rely on users or administrators opening an e-mail attachment and/or rendering an HTML formatted message, which replaces a valid element type with a disguised malicious code. A Malicious Code attack may also be an end in itself, intended to bring havoc and shut down a server or network, working similarly to a DoS attack.

USER COMPROMISE

This variable represents the damage to the user's system that happens in the later stages of Step 2, after the door to a user's computer is opened through malicious code, brute force, or physical access. While Malicious Code is a measure of attempts to compromise computer systems, User Compromise is a measure of actual breaches of user computer systems.

ROOT COMPROMISE

Root Compromise corresponds to Step 3 in the hacking framework. Once hackers succeed at User Compromise, they try to gain administrator or root-level privileges (escalation of privilege), and to consolidate power by obtaining other accounts, and accessing other resources. Root Compromise may also be achieved directly, through Malicious Code that takes advantage of operating systems, server and application vulnerabilities, typically using buffer overflows to deliver the malicious payload. Root Compromise and Malicious Code are defined as mutually exclusive categories in the FCIRC data, so there is no double counting in this study.

DENIAL OF SERVICE

Denial-of-Service attacks (Section II) represent attempts to make a service inaccessible to legitimate users, as happened in the notorious attacks against Yahoo, Microsoft and others. When hackers are not able to achieve Root Compromise, they often express their frustration by launching DoS attacks, where they flood a network or disrupt connections or services [McClure et al., 2001; Panko, 2003; Pfleeger and Pfleeger, 2002]. The assumption here is that if hackers cannot achieve control of the resource, they will try to at least make it inaccessible.

HOSTS AND BROADBAND

Several environmental factors might contribute to the growth in security threats. Given the exploratory nature of this study, we consider only two such environmental variables: number of hosts in the Internet (Hosts) and broadband access for households and small business (Broadband). The increase in the number of hosts can be seen as an increase in the number of potential targets for hacker attacks [Pfleeger and Pfleeger, 2002]. The increase in broadband Internet access for households and small businesses means that there are now millions of computers working on 24 hours, 7 days a week, with little or no protection. Because households and small businesses typically lack the security protections used by large businesses, anecdotal evidence indicates that the increase in broadband DSL and cable users is associated with the increase of malicious code attacks [McClure et al., 2001].

IV. RESEARCH QUESTIONS AND HYPOTHESES

This exploratory study addresses two broad research questions:

- Question 1. Which of the relationships presented in the model can be observed in practice?
- Question 2. Are environmental variables such as number of hosts and broadband access associated with the hacking activities presented in the model?

These research questions led to the formulation of the following hypotheses:

Hypothesis 1: **Reconnaissance is positively related to Broadband and Hosts**

The increase in broadband access and in the number of Internet hosts represents an increase the number of potential hacker targets. Faced with more targets, hackers may increase their information gathering activities for target acquisition (i.e., Reconnaissance).

Hypothesis 2: **Malicious Code is positively related to Reconnaissance**.

An increase in Reconnaissance activities may lead to finding a higher number of potential victims, which in turn may increase hacker attempts to distribute Malicious Code.

Hypothesis 3: **User Compromise is positively related to Broadband, Reconnaissance, and Malicious Code**.

The increase in Broadband access may represent an increase the number of users with lower computer security in place. This increased vulnerability may contribute to an increase in User Compromise.

An increase in Reconnaissance activities may lead to the identification of more potential desirable targets, which may then lead to an increase in initial access and contribute to User Compromise.

An increase in Malicious Code attacks may lead to opening doors into an increased number of user systems, resulting in an increase in User Compromise.

Hypothesis 4: **Root Compromise is positively related to Hosts, Malicious Code, and User Compromise**.

The increase in the number of Internet Hosts represents an increase in the number of computer systems with administrator or root privileges, which in turn may increase the opportunities for Root Compromise.

Malicious Code attacks may enable hackers to achieve administrator or root-level privileges, thus contributing directly to an increase in the number of successful incidents of Root Compromise.

Malicious Code attacks may also contribute to Root Compromise through a more indirect route, where hackers first gain control over a user's system (User Compromise) and then, through privilege escalation, achieve Root Compromise.

Hypothesis 5: **Denial of Service is positively related to Hosts and Malicious Code, and negatively related to Root Compromise.**

The increase in the number of Internet Hosts, and the corresponding increase in the number of servers, may represent an increase in the opportunity for DoS attacks.

An increase in Malicious Code may also lead to increase in DoS attacks (e.g., the SCO Denial of Service attack caused by *MyDoom*).

When hackers achieve little or no Root Compromise, frustration may lead to an increase in the number of DoS attacks.

Hypothesis 6: **Broadband is positively related to Hosts**.

The growth of the Internet (reflected in the growth in Number of Hosts) can be expected to increase the demand and availability of broadband access.

V. MEASUREMENT AND DATA COLLECTION

Most studies of security breaches rely on self-reported incidents, identified through the use of samples and surveys. In this study, we decided to explore the possibility of using non-reactive measures, based on available official descriptive statistics of the whole population, rather than samples. We used 60 months (five years) of statistical data on incidents reported from 1999 (when the federal government started collecting those statistics) through 2003. Given the exploratory nature of the study, we judged that this amount of accumulated data was large enough to offer insights into the methodological promise of using non-reactive data about the whole population. The continuing accumulation of new data over the next several years will generate longer historical series and allow more sophisticated analyses in the future (e.g., testing for the possibility of cyclical phenomena that peak at certain times of the year).

The variables (Reconnaissance, Malicious Code, User Compromise, Root Compromise, and Denial of Service) were measured using the data collected for the similarly named categories in the statistics published by the Federal Computer Incident Response Center between 1999 and 2003 [FCIRC, 2004].

We found no reliable worldwide statistics for the number of Hosts in the Internet. Therefore, in this exploratory study, we used a surrogate measure, the number of web servers on the Internet, which reflects the expansion in the use of the web and access to the Internet. Our source for the total number of Internet servers from 1999 to 2003 was Netcraft, where this statistic is gathered by querying all servers in the Internet [Netcraft, 2004]. We used the total number of servers because we were interested in the overall expansion or growth curve, not in the relative numbers of different types of web servers.

The Broadband variable in the model was measured using data on the 60-month growth of broadband access for household and small businesses, which we obtained from statistics collected by the FCC between 1999 and 2003 [Federal Communications Commission, 2004].

VI. ANALYSIS AND RESULTS

DESCRIPTIVE STATISTICS

Table 1 shows the descriptive statistics for the variables in the model. Reconnaissance and Malicious Code account for 97% of the incidents reported. The statistics in Table 1 are particularly important because of their impact on such a large number of users and computers. Delivery of viruses to mail users by Malicious Code (as in the case of *Melissa*), can lead to the shutdown of mail servers, thus hurting not only the users who are actually infected by the viruses, but also all other users who can no longer access the mail server. Once a server is unavailable (as in the case of the Denial of Service attack at Microsoft), an incalculable number of users can be affected by such factors as the inability to obtain information or download latest updates and patches.

Table 1. Descriptive Statistics

Variable	Mean/Month	Total	% of Attacks
Reconnaissance	19,920	1,195,186	85.74
Malicious Code	3,272	196,335	14.08
User compromise	16	961	.07
Root compromise	11	633	.04
Denial of Service	15	891	.06
Total incidents(*)	23,233	1,394,006	

*Note: This total number of incidents does not include unclassified incidents

Although User and Root Compromise are only achieved in a small percentage of all attacks, the results in Table 1 mean that significant security breaches still occur almost every day. As discussed in Sections II, III and IV, Root Compromise can be achieved directly or indirectly, through User Compromise and escalation of privilege. Regardless of how it is accomplished, each incident of Root Compromise potentially can affect a large number of computers in an organization, given that hackers who achieve root or administrator level privileges can then access most of the other computers in a local area network.

SIMPLE CORRELATIONS

Table 2 shows the simple correlations of all the variables in the model. The significant correlations appear in bold, and the letters in parentheses indicate their level of significance. Some high correlations were also highly significant (e.g., Broadband and Hosts), while other correlations were low and with low significance (e.g., Denial of Service and Hosts).

Table 2. Simple Correlations

	Hosts	Broadband	Recon-naissance	Malicious Code	User Compromise	Root Compromise	Denial of Service
Hosts	1	**.9022** (a)	**.4195** (b)	**.3240** (c)	**.1809** (e)	-.0297	**.1505** (e)
Broadband	**.9022** (a)	1	**.5135** (a)	**.4015** (b)	**.2223** (d)	-.0347	.0021
Recon-naissance	**.4195** (b)	**.5135** (a)	1	**.6230** (a)	.0025	.0025	-.1109
Malicious Code	**.3240** (c)	**.4015** (b)	**.6230** (a)	1	.1090	-.1142	-.0432
User Compromise	.1809	**.2223** (d)	.0025	.1090	1	**.1743** (e)	-.0630
Root Compromise	-.0297	-.0347	.0256	-.1142	**.1743** (e)	1	**-.2980** (e)
Denial of Service	**.1505** (e)	.0021	-.1109	-.0432	-.0630	**-.2980** (e)	1

(a) significant at .0001 (b) significant at .001 (c) significant at .01 (d) significant at .1 (e)significant at .2

HYPOTHESES TESTING

Table 3 shows the results of the stepwise regression used to test the hypotheses listed in Section IV.

Table 3. Results of Stepwise Regression

	Dependent Variables	Independent Variables	BETA	t significance	R^2	Hypothesis Support?
H1	Reconnaissance	Broadband	.513476	.00001	.26366	Partial
H2	Malicious Code	Reconnaissance	.622971	.00001	.38809	Yes
H3	User Compromise	Broadband	.222302	.0878	.04942	Partial
H4	Root Compromise	User Compromise	.174308	.1829	.03038	Partial
H5	Denial of Service	Root Compromise Hosts	-.293751 .141746	.0224 .2619	.10886	Yes
H6	Broadband	Hosts	.902162	.00001	.81390	Yes

Hypothesis 1 is partially supported. Reconnaissance is positively related to the growth in Broadband access, and Broadband variation explains almost 30% of Reconnaissance variation. The association between Reconnaissance and Hosts, however, is not significant.

Hypothesis 2 is supported. Malicious Code is positively related to the increase in Reconnaissance, and Reconnaissance variation explains almost 40% of Malicious Code variation.

Hypothesis 3 is partially supported. User Compromise is positively related to Broadband, probably due to unprotected user systems in homes and small businesses. Surprisingly, User Compromise is not related to increase in Reconnaissance or Malicious Code. Given that only about 5% of User Compromise is explained by Broadband variation, it seems that other factors not considered in the model may have a greater influence on User Compromise (e.g. the number of vulnerabilities found in popular operating systems and applications).

Hypothesis 4 is partially supported. Root Compromise is modestly associated with User Compromise (3%), probably through escalation from user to root privilege. Malicious Code and Hosts do not affect Root Compromise, however. Again, it seems that other factors not considered in the model may have a greater influence on Root Compromise.

Hypothesis 5 is supported. Denial of Service is positively related to Hosts and negatively related to Root Compromise, as expected. Given that Root Compromise and Hosts explain only about 10% of DoS variation, other variables that were not included in the model should contribute to Denial of Service. Nevertheless, the results are still compatible with the idea that when hackers are frustrated in their efforts to compromise a system, they may try instead to make the resource inaccessible.

Hypothesis 6 is supported. The results show a strong positive relationship between the environmental variables used in the model (Broadband and Hosts). More than 80% of Broadband variation is explained by Host variation.

Figure 2 presents the relationships of the variables in the model, with the BETA values obtained for the relationships that were found to be significant. Once more we should point out that this study is too exploratory in nature to allow a full causal path analysis, and therefore no causality should be inferred from these results.

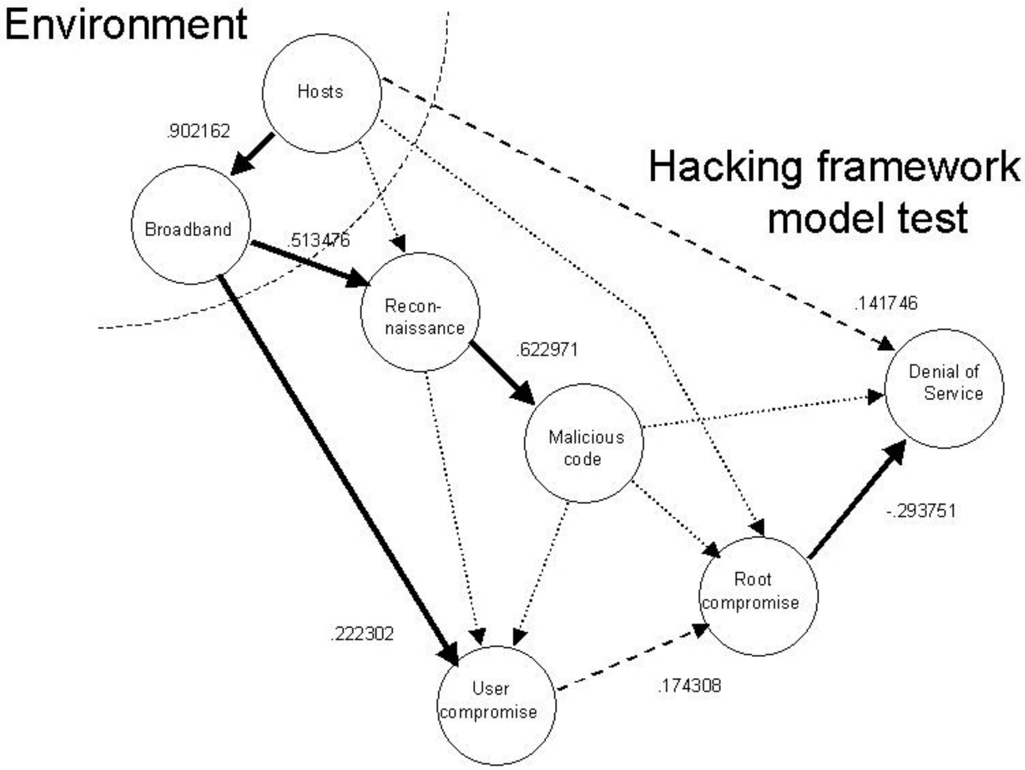

Figure 2. Test of Model

VII. CONCLUSION

The results of this exploratory study provide preliminary support for the model inspired by the Hacking framework. As expected, Broadband Internet access by home and small business users is positively related to Reconnaissance activities, suggesting a relationship between the number of vulnerable systems and the attempts of hackers to find them. Reconnaissance (Step 1) is positively related to Malicious Code (the preferred route to achieve initial access in Step 2), supporting the model's prediction that increased efforts to find vulnerable systems are associated with increased attempts to break into them. Success in achieving User Compromise in Step 2 is positively related to Root Compromise (Step 3), which is compatible with hackers' attempts towards escalation of privilege. The finding of a negative relationship between Root Compromise and Denial of Service supports the idea that a hacker's frustration at failing to gain control of a resource may be associated with trying to sabotage it through DoS attacks. However, it seems that the model still lacks other variables that might further explain User Compromise, Root Compromise and Denial of Service, and more research is needed to identify and test those missing variables.

The exploratory nature of the study only allows for suggestive, rather than prescriptive implications for practice. Some practical implications, however, deserve special note. For example, the positive relationship between Reconnaissance and Malicious Code suggests that Systems Administrators should consider the detection of Reconnaissance activities as an early signal of future attacks using Malicious Code. This early signal should allow systems administrators to be prepared and take preventive measures to strengthen their defenses ahead of such Malicious Code attacks.

System Administrators should also be aware of the potential risk represented by employees who have broadband connections at home and use them to gain remote access to the company's network. Although the relationship between the two variables was not very strong, the results suggest that User Compromise at home may lead to Root Compromise at work.

Broadband Internet service providers should note the positive relationship between User Compromise and increase in broadband access. An important practical implication of this finding is that service providers should try to prevent User Compromise by creating network security mechanisms that compensate for the typical home user's lack of training in computer security.

The exploratory nature of this study leaves ample room for further development and testing of the model. Future research should use more rigorous methods, such as causal path analysis, to test for causality. Future studies should also test for serial correlations (given the historical nature of the data), for multi-colinearity (given a high level of correlation between the environmental variables used in the model), and for non-linearity in the relationships between some of the variables (given the ever- accelerating rate of Internet growth).

Despite the constraints of this exploratory study, we believe that one of its significant contributions is the attempt to use secondary data, based on official statistics, which are non-reactive and collected at the time of the security incidents. Most of the existing studies of security breaches rely on data from surveys where respondents in a sample are asked to report, retroactively, their experience with hacking attacks. We believe, however, that there is significant promise in data collection strategies such as the one used here, based on using secondary data obtained from publicly available sources which collect longitudinal series of statistics for the whole population under study.

For example, future researchers trying to identify additional environmental variables to explain User and Root Compromise may find it helpful to use archives from Microsoft, RedHat, Sun, Symantec, and other sources. This archival data should allow researchers to identify the frequency and nature of discovered vulnerabilities, and the number and security threat levels of the malicious codes created to exploit these vulnerabilities. Using such industry statistics, researchers should then be able to expand our understanding of the relationships between successful Root or User Compromise and the number of computers using different operating systems, servers and applications.

As government and other research and industry centers continue to collect data over the next several years, it should be possible to examine questions such as the possibly cyclical nature of hacking attacks (for example, whether they peak at certain times of the year, such as the holiday shopping season). The ongoing accumulation of statistical data will make possible the analysis of much longer historical series and allow a deeper understanding of the phenomena in the hacking framework, hopefully enhancing our ability to prevent and reduce security breaches.

ACKNOWLEDGEMENTS

The authors acknowledge the helpful input from reviewers and participants in the 2004 Americas Conference in Information Systems.

Editor's Note: This article was received on October 13, 2004 and was published on December __, 2004

REFERENCES

> *Editor's Note:* The following reference list contains hyperlinks to World Wide Web pages. Readers who have the ability to access the Web directly from their word processor or are reading the paper on the Web, can gain direct access to these linked references. Readers are warned, however, that
> 1. these links existed as of the date of publication but are not guaranteed to be working thereafter.

2. the contents of Web pages may change over time. Where version information is provided in the References, different versions may not contain the information or the conclusions referenced.
3. the author(s) of the Web pages, not AIS, is (are) responsible for the accuracy of their content.
4. the authors of this article, not AIS, are responsible for the accuracy of the URL and version information.

Bagchi, K. and G. Udo (2003) "An Analysis of the Growth of Computer and Internet Security Breaches," *Communications of the Association for Information Systems* (12), pp. 684-700.

Bento, A. (2003) "Soho Security: A Technical Briefing," *Proceedings of the Americas Conference on Information Systems.* Tampa, Florida.

Bookholdt, J. L. (1989) "Implementing Security and Integrity in Micro-Mainframe Networks," *MIS Quarterly* (13) 2, pp. 135-144.

CERT/CC (2004) "CERT/CC Statistics 1988-2004" http://www.cert.org/stats/ cert_stats.html (current September 29, 2004)

CERT / CC (2002). "Overview of Attack Trends" http://www.cert.org/archive/pdf/ attack_trends.pdf (current September 29, 2004)

Federal Communications Commission (2004) "Local Telephone Competition and Broadband Deployment"http://www.fcc.gov/wcb/iatd/comp.html (current September 29, 2004)

Federal Computer Incident Response Center (2004) "U.S. CERT Federal Incident Statistics" http://www.us-cert.gov/federal/statistics/ (current September 29, 2004)

Garg, A., J. Curtis and H. Halper (2003) "Quantifying the Financial Impact of IT Security Breaches," *Information Management & Security* (11) 2, pp. 74-83.

Howard, J. D. (1997) *An Analysis of Security Incidents on the Internet*. Ph.D. Dissertation, Pittsburgh, PA: Carnegie Mellon University.

Howard, J. D. and T. A. Longstaff (1998) *A Common Language for Computer Security Incidents*. Albuquerque, NM: Sandia.

Krol, E. and P. Ferguson (1995) *The Whole Internet*. Sebastopol, CA: O'Reilly.

Landwehr, C. E. et al. (1994) "A Taxonomy of Computer Program Security Flaws," *ACM Computing Surveys* (26) 3, pp. 211-233.

Loch, K.D., H. H. Carr, and M. E. Warkentin (1992) "Threats to Information Systems," *MIS Quarterly* (16) 2, pp. 173-186.

Lukasik, S. J. (2000) "Protecting the Global Information Commons." *Telecommunication Policy* (24), pp. 519-531.

McClure, S., J. Scambray and G. Kurtz (2001) *Hacking Exposed*. New York, NY: McGraw-Hill.

Netcraft (2004) "Web Server Survey" http://news.netcraft.com/archives/ web_server_ survey.html (current September 29, 2004)

Panko, R. (2003) *Business Data Networks and Telecommunications* Upper Saddle River, NJ: Prentice-Hall.

Pfleeger, C, P, and S. L. Pfleeger (2002) *Security in Computing* Upper Saddle River, NJ: Prentice-Hall

Richardson, R. (2003) *The 2003 CSI/FBI Computer Crime and Security Survey*. San Francisco, CA: Computer Security Institute.

Straub, D.W. and W. D. Nance (1990) "Discovering and Disciplining Computer Abuse in Organizations," *MIS Quarterly* (14) 1, pp.45-60.

Straub,D.W. and R. J. Welke (1998) "Coping with Systems Risk," *MIS Quarterly* (22) 4, pp. 441-469.

ABOUT THE AUTHORS

Al Bento is Professor of Information Systems at the Merrick School of Business, University of Baltimore. He is Editor-in-Chief of *the Journal of Information Technology Management* and Chair of the Association for Information Systems (AIS) Special Interest Group on Information Systems Security (SIGSEC). A graduate from UCLA, his current research includes projects on information security and performance measurement systems.

Regina Bento is Professor of Management and holder of the Hatfield-Merrick Distinguished Professor chair at the Merrick School of Business, University of Baltimore. She is past Chair of the Management Education and Development Division at the Academy of Management, and "OB-1" at the Organizational Behavior Teaching Society. A graduate from MIT, her current research includes projects on the behavioral aspects of the use of information technology.

Communications of the Association for Information Systems

ISSN: 1529-3181