

Drivers of Information Security Search Behavior: An Investigation of Network Attacks and Vulnerability Disclosures

JINGGUO WANG

The University of Texas at Arlington

and

NAN XIAO and H. RAGHAV RAO

University at Buffalo

More and more people use search engines to seek for various information. This study investigates the search behavior that drives the search for information security knowledge via a search engine. Based on theories in information search and information security behavior we examine the effects of network attacks and vulnerability disclosures on search for information security knowledge by ordinary users. We construct a unique dataset from publicly available sources, and use a dynamic regression model to test the hypotheses empirically. We find that network attacks of current day and one day prior significantly impact the search, while vulnerability disclosure does not significantly affect the search. Implications of the study are discussed.

Categories and Subject Descriptors: H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Search process*

General Terms: Security, Management, Human Factors

Additional Key Words and Phrases: Information systems, information security, information search behavior, search engine, network attacks, vulnerability disclosures, dynamic regression

An early version of this article was presented at the 8th Workshop on eBusiness (WeB'09).

The work of J. Wang was supported by a summer research grant from UT Arlington College of Business. The research of H. R. Rao was funded in part by NSF under grant 0916612 and by Sogang Business School's World Class University Project (R31-20002) funded by Korea Research Foundation.

H. R. Rao is SUNY Distinguished Service Professor, SUNY Buffalo and WCU Visiting Professor, Sogang University. He is also affiliated with the Department of Service Systems Management & Engineering, Sogang University.

Authors' addresses: J. Wang, Department of Information Systems and Operations Management, College of Business, the University of Texas at Arlington, Arlington, TX 76019; email: jwang@uta.edu; N. Xiao and H. R. Rao, Department of Management Science and Systems, School of Management, University at Buffalo, Buffalo, NY 14260; email: {nanxiao, mgmtrao}@buffalo.edu.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.

© 2010 ACM 2158-656X/2010/12-ART3 \$10.00

DOI 10.1145/1877725.1877728 <http://doi.acm.org/10.1145/1877725.1877728>

ACM Transactions on Management, Information Systems, Vol. 1, No. 1, Article 3, Publication date: December 2010.

ACM Reference Format:

Wang, J., Xiao, N., and Rao, H. R. 2010. Drivers of information security search behavior: An investigation of network attacks and vulnerability disclosures. *ACM Trans. Manag. Inform. Syst.* 1, 1, Article 3 (December 2010), 23 pages.
DOI = 10.1145/1877725.1877728 <http://doi.acm.org/10.1145/1877725.1877728>

1. INTRODUCTION

Humans may be viewed as Informavores [Miller 1983]. Information needs arise when there is a gap between the contextual situation and the desired situation (e.g., uncertainty) [Dervin 1983]. Informavores, in the online context, satisfy their information needs by searching on the Web. The Web has huge amounts of information, and search engines have become the primary, powerful, and convenient tool for users to search for and find relevant information. Searching is a means to recover what a seeker knows exists on the Web and also to discover what a seeker assumes must be on the Web [Battelle 2005]. Users may perform online searches by formulating and submitting queries to search engines [Venkatsubramanyan and Kwan 2008]. In fact, millions of individuals regularly use modern search engines for topics ranging from news to health-care. More than 80% of Internet users use search engines as their starting point of Web surfing [Jansen 2007].

Several studies have investigated the motives of online information search, and showed that users' concerns about their well-being (such as health and employment) can lead to search. Using U.S. government monthly unemployment reports and search tracking data, Ettredge et al. [2005] found the frequency of job-related search is positively associated with the number of unemployed for each month. Studies on health-information-seeking behavior and health-care concerns have yielded similar results and revealed that if people perceive they are susceptible to diseases, they are motivated to seek health information. Ginsberg et al. [2008] showed high correlations between the frequency of influenza-related search and the percentage of physician visits in which a patient presents influenza-like symptoms.

Analyzing Web searches does not just reflect what is happening online but gives a wider picture of society and people's behavior [Tancer 2008]. Such analyses throw light on how people are navigating the information universe and have been used by Google to create the multibillion dollar industry of paid search [Battelle 2005] (e.g., AdSense and Adwords). Often companies have analyzed clickstream patterns to deliver innovative Web-based services that are better and useful to the navigators [Battelle 2005]. For instance, based on the fact that clickstream patterns changed after 9/11, Google created a new service: Google News [Battelle 2005]. Battelle [2005] pointed out that according to Piper Jaffray, 65% of Internet searches are informational, but the average number of queries per visit to an engine is nearly five; the implication is that information seekers are not getting what they want the first time or have to come up with new queries driven by the results of the initial query. This shows that Web search is obviously not as effective as it could be. This has critical implications because as Rose and Levinson [2004] point out, searching

is a means to an end “to satisfy the underlying goal that a user is trying to achieve.”

In this study we focus on one important domain: information security. As information security has become a major concern for Internet users, a large body of knowledge on the topic has been created and accumulated online. Because ordinary users (i.e., Internet users without much knowledge and experience regarding information security) may lack the experience and knowledge required to deal with information security, their information needs arise. Search engines relieve users and can locate needed information rapidly. Not only can such information be used to protect a computer or recover a computer from virus afflictions, but also to help to understand the characteristics of security threats and determine whether a particular threat is relevant or not. Additionally, information security-related searches may be a precursor to protective behavior (e.g., periodically updating operating systems and antivirus software, installing a firewall, not opening suspicious emails from unknown sources, and securing passwords [Anderson and Agarwal 2010] as well as using intrusion detection systems and access control mechanisms). This is analogous to research in healthcare that showed that the act of exploring for illness information can reduce users’ anxiety, increase feelings of self-efficacy, improve the effectiveness of protection [Ybarra and Sumanb 2006], and trigger the exercise of self-care [Chou and Wister 2005].¹ Thus it is important to explore and understand the characteristics of information security search behavior.

When affected by an information security threat, the uncertainty that ordinary users face is high. For example, before taking proper protective action, users need to assess the severity and relevance of the information security threat and evaluate the effectiveness of different protective methods. To reduce the uncertainties and make informed decisions, they tend to collect sufficient information on a variety of aspects of the information security threat. In addition, information seeking could affect peoples’ perception of risk, efficacy, and attitude towards protection acts [Griffin et al. 1999; Severtson et al. 2006]. Therefore, information seeking is an important element of information security protective behavior and also directly influences other subsequent elements of protective behavior. This has been shown to be the case in other related arenas, such as mass media reports and hazards [Neuwirth et al. 2000] as well as search for radon risk information that increases the likelihood of risk mitigating and protective actions [Smith et al. 1995]. As such, one can design message interventions that can help in developing proper preventive behaviors [Griffin et al. 1999]. From a business standpoint, Web sites that provide information security-related information would be interested in search behavior because it drives ad revenue (if their site contains ads).

This study investigates the drivers of search for information security knowledge via a search engine by ordinary users. Using the AOL search log (captured between March 2006 and May 2006),² we empirically examine the effects of network attacks and vulnerability disclosures on the search behavior of ordinary

¹We thank an anonymous referee for the suggestion.

²The three-month search log is the dataset released by AOL Research in August 2006.

users for information security knowledge. We organize the article as follows. Section 2 develops our hypotheses, drawing from theories in information search and information security behavior. Section 3 introduces the data and measurement, and presents the empirical modeling approach and regression results. Section 4 discusses the implications of the study.

2. THEORETICAL BACKGROUND

In this section, we draw from theories in information search and information security behavior to develop conjectures regarding the relationship between network attacks and search for information security knowledge, and that between vulnerability disclosures and search for information security knowledge. These conjectures are used to guide us in the exploration of empirical regularities in search behavior as it pertains to network attacks.³ A number of studies from different disciplines such as physiology, communication, and healthcare have examined the motivation and strategy of information search behavior. In this study, we view information seeking about security knowledge as protection seeking and protective actions, which occurs as part of both threat appraisal and threat coping processes [Brouwers and Sorrentino 1993; Eppright et al. 1994; Rippetoe and Rogers 1987; Srinivasan and Ratchford 1991], though the kind of information sought in these two processes may differ [Neuwirth et al. 2000]. In a threat appraisal process because an individual's information-processing goal is to determine whether a particular security threat is relevant, information about the threat's characteristics (severity, threat likelihood) would be the most related. But in a threat coping process with a goal of controlling the threat, the information sought would focus on a remedial measure's effectiveness and to what extent a person could expect to successfully carry out a corrective action. Prior studies have also provided empirical evidences in supporting protection motivation theory [Rogers 1975] to explain general information seeking behavior about a hazard (without distinguishing the kind of information sought) [Anderson and Agarwal 2006; Brouwers and Sorrentino 1993; Eppright et al. 1994; Herath and Rao 2009b; Neuwirth et al. 2000; Rippetoe and Rogers 1987; Srinivasan and Ratchford 1991; Woon et al. 2005; Workman et al. 2008].

2.1 Network Attacks

While computer technology and Internet connections have fundamentally changed our way of life, various attacks over networks pose severe threats. A report by the Computer Security Institute [Richardson 2008b] showed that 32% of reporting organizations have been attacked by malware in 2008. As there is always a portion of computer users that do not install proper security measures such as antivirus software and firewalls, or fail to update their operating systems and/or security measures, their systems are vulnerable to network attacks. A computer/network breach may cause serious losses, including loss of

³We thank an anonymous referee for this observation.

confidential data and system failure. Users' behavior of search for information security knowledge can be affected by network attacks directly or indirectly. Those users whose systems have been infected by viruses or compromised by a hacker will have to take prompt actions to identify and diagnose problems, recover the system, and mitigate the damage.

Even if a user is not directly affected by network attacks, he/she may be aware of such ongoing attacks from media coverage or individuals experiencing attacks. Network attacks, especially the serious ones (e.g., Melissa, Love Bug, and Code Red), usually draw media's attention and are covered rapidly. Media coverage on network attacks can alter users' risk perception [Rogers 1997]. The relationship between news coverage and subsequent information search has also been documented, and Cooper et al. [2005] found that news coverage on cancer generates users' search for information on cancer. Although mass media plays an important role in improving public awareness and shaping public attitude towards information security, media reports on network attacks tend to emphasize the consequence and lack the details on how to protect computer systems [Dowland et al. 1999]. Thus, users who are influenced by the reports have to turn to other sources for additional information which helps determine the relevance and/or provides protective approaches. Besides, users' perceived risk of network attack may also be influenced by their peers (such as friends, colleagues) experiencing attacks [Herath and Rao 2009a]. Consequently they may be more alert.

Network attack could directly and indirectly increase users' perceived risk, which in turn could trigger users' protection behaviors [Kumar et al. 2008; Liang and Xue 2009; Ng et al. 2009]. To avoid or minimize their losses, users will actively seek approaches to protect their systems when perceiving the immediate attack threat is high. Protecting the system from attack requires sufficient knowledge and proper tools; however, most individuals, including some professionals, lack such knowledge and technology solutions to protect them against these attacks [Furnell et al. 2006]. Therefore, they may rely on search engines to find related solutions or suggestions once faced with the attacks.

Therefore, we propose that when network attacks intensify, in order to prevent the afflictions and reduce the damages associated with the attacks, users will be actively engaged in seeking information security-related knowledge via search engines. Network attacks can be measured from two dimensions: intensity and prevalence. Attack intensity has been considered as an important factor affecting the effectiveness of intrusion detection [Manikopoulos and Papavassiliou 2002] and classifying denial-of-service attacks [Hussain et al. 2003]. Attack prevalence has been used by antivirus software vendors (e.g., <http://www.virusbtn.com>) to measure the risk of malicious codes and investigate the propagation of computer virus [Kephart and White 1993]. The intensity of network attacks refers to the average number of attacks arriving at a computer in a unit time. High-intensity attacks increase the possibility of a system being breached. The prevalence of network attacks refers to the number of computer systems targeted by the attacks. Since some computers may not be properly protected, a more prevalent attack is more likely to affect a larger number of computer users. Therefore we propose the following hypotheses.

H1(a). The intensity of network attacks increases the frequency of search for information security knowledge.

H1(b). The prevalence of network attacks increases the frequency of search for information security knowledge.

Since system disruption can interrupt users' normal computer usage or even their lives and work, recovery or prevention is urgent. Users may respond quickly to attacks. For Internet users, initiating a search via search engines does not demand sophisticated skills. Thus it is possible that the search for information security may take place shortly after the occurrence of network attacks. There are two possible outcomes associated with a search for information security. First, users may find information they need and consequently complete the search [Browne et al. 2007]. Second, users may not find useful information online after trying different queries. Due to the urgency of the situation, they may cease searching and turn to other channels, such as asking for help from friends, or contacting software vendors. Search for information security information is users' emergency response to network attacks. Once their information needs related to the network attacks are satisfied (normally within a short period, e.g., within a few days, maybe from different channels), users may stop their search on the topic. As such, it is likely to be driven by network attacks. Search for information security may diminish as the concerns regarding networks are no longer valid. Unlike the search triggered by users' interests and enjoyment or those directly related to users' daily life which may last over an extensive period of time, search for information security may only occur in a short period of time when the concerns on network attacks are still valid. Therefore we propose our next hypotheses.

H2(a). The intensity of network attacks has immediate but short-term impacts on the frequency of search for information security knowledge.

H2(b). The prevalence of network attacks has immediate but short-term impacts on the frequency of search for information security knowledge.

2.2 Vulnerability Disclosures

A vulnerability is a technical flaw or weakness in the design, implementation, or operation and management that can be exploited to violate a system's security policies [Png et al. 2008]. Most of vulnerabilities originate from software vendors' design flaws. The vulnerabilities pose potential risks to users, and such risks are realized when they are exploited by network attacks.

Vulnerability disclosures refer to making vulnerability information public. A vulnerability could either be disclosed immediately with full details after it is discovered, or be disclosed after a certain period of time with limited details to allow the vendors to develop and release the patch [Li and Rao 2007]. Vulnerability disclosures have conflicting effects. On one hand, vulnerability disclosures could press the vendor to release a patch sooner and help users take proactive actions [Arora et al. 2010; Cavusoglu et al. 2007]. On the other hand,

attackers could take advantage of the disclosed information to launch attacks and threaten systems [Png et al. 2008].

The number of vulnerability disclosures reflects the amount of flaws that exist in a variety of software packages. A large number of vulnerability disclosures may affect a larger number of users, and thus raise more users' concerns about the security of their systems. The severity of a vulnerability represents the impact of vulnerability on IT assets, the remediation status of the vulnerability, and the availability of exploit code [Mell et al. 2007]. More severe vulnerabilities may make users worry more about their information security. According to Protection Motivation Theory (PMT) [Rogers 1975], these concerns would result in users' intention to take actions to protect their systems, which in turn affects the frequency of users' search for information security knowledge. Further, ordinary users would have trouble differentiating between a vulnerability disclosure and an actual attack, and consequently vulnerability disclosures would have a positive influence on information search behavior.

However, vulnerability disclosures may have limited impacts on users' search for information security knowledge due to the following four reasons. First, the communication channel used by vulnerability disclosures limits the information dissemination to general public. Vulnerability disclosures are aimed at vendors and information security professionals rather than ordinary users, and such information is normally found at only a few information-security-dedicated Web sites. Ordinary users may be unaware of such disclosures as they do not check these Web sites so regularly. While some Web sites send vulnerability information to users by email, this service is usually limited to subscribers. Further, vulnerabilities can exist for a very long time without ever being addressed. McQueen et al. [2009] showed that many vulnerabilities in the National Vulnerability Database remain dormant and unaddressed, sometimes for years, until an exploit occurs. The value of the vulnerability information channel as a driver of end-user behavior is therefore less than exploit information. This supports the argument that vulnerability information itself is of interest to vendors rather than end-users.⁴ Second, most vulnerability announcements contain many technical terms. Usually, they are brief and the details have been hidden intentionally. Hence, for ordinary users, it is not straightforward to know how these vulnerabilities could affect their systems. The lack of capability could reduce the motivation of search for information security knowledge [Rogers 1975]. Third, as suggested by protection motivation theory, response cost, which refers to any cost associated with taking protection actions, will reduce the possibility of response [Floyd et al. 2000]. Since search for information about a specific vulnerability is not easy, the associated search cost could be higher. Therefore, despite awareness of vulnerability disclosures, it is possible that the users tend to ignore it and avoid search costs. Fourth, based on threat avoidance theory, only after a threat is perceived is the user motivated to search for information related to coping with the threat [Liang and Xue 2009]. Vulnerabilities, unlike network attacks, do not cause immediate damage until they are exploited. Consequently, users may be less

⁴We thank an anonymous referee for pointing this out.

motivated in seeking for information security knowledge when vulnerabilities are disclosed.

Therefore, we propose the following alternate pairs of exploratory hypotheses.

H3(a1). The number of vulnerability disclosures increases the frequency of search for information security knowledge.

H3(b1). The severity of disclosed vulnerability increases the frequency of search for information security knowledge.

H3(a2). The number of vulnerability disclosures does not increase the frequency of search for information security knowledge.

H3(b2). The severity of disclosed vulnerability does not increase the frequency of search for information security knowledge.

3. EMPIRICAL ANALYSIS

In this section, we first introduce our data and measurements. We then discuss our empirical modeling approach, and present the estimation results.

3.1 Data and Measurements

Search engine logs have been widely used in previous literature to study users' sociological behavior and information needs [Chau et al. 2007; Richardson 2008a]. We derive the frequency of search for information security knowledge using a search log from AOL Research [Pass et al. 2006]. AOL⁵ is known to be one of the big four portals that most searchers use (the others are Microsoft, Yahoo, and Google [Battelle 2005]). According to comScore Media, as of July 2006, the percentage of searches done by U.S. Web surfers at home, work, and at universities that were performed at AOL is 5.9%.⁶ According to Alexa.com, compared with the general internet population, AOL has a higher percentage of users over age 45 and between 25 and 34. It also has a higher percentage of female users. The browsing location of a higher percentage of AOL users is at home. In addition AOL has a higher percentage of users with some college education, and less with graduate school.

The search log we use includes about 658000 users over a three-month period from March 1 to May 31, 2006. It covers almost 1.5% of the May search users at AOL according to ComScore Media Metrix. The log has about 20 million search records, and around 1/3 of 1% of the total searches conducted through the AOL network over that period. The search log only includes U.S. searches conducted within the AOL client software. We use keyword matching to identify the queries that were intended to seek for information security-related topics. The list of keywords were generated based on the CSI/FBI computer crime and security survey [Richardson 2008b] and suggestions from

⁵Note: In May 2002, AOL and Google inked a landmark deal wherein Google provided search technology to AOL's network of Internet properties [Battele 2005].

⁶see <http://searchenginewatch.com/2156431>.

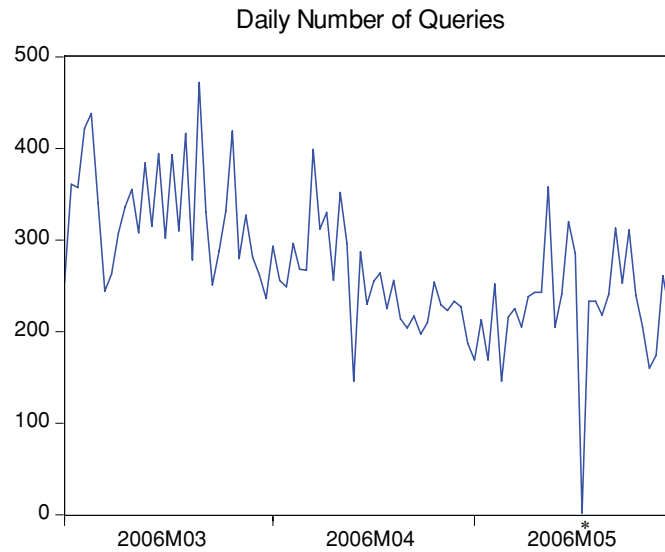
information security professionals. We also recruited a class of masters students in information systems (44 students in total) and asked them to list common queries that might be used to search for information security knowledge via a search engine. Thirty-two of these students had bachelor degrees in information technology or computer science. They had an average of 25-months work experience in either an IT company or the IT department of a traditional company. The complete list of the keywords we used is listed in Table I. The daily number of the queries (denoted as $Y_t, t = 1, \dots, 92$) that contain at least one of these keywords are used to measure the frequency of search for information security knowledge. Figure 1 plots the series. We correlated the series with the search volume from Google Trends in that period, and have a correlation coefficient of .50 (significant at 0.001). Note that due to the restriction of the number of keywords that we are able to input in the search box of Google Trends, we grouped these keywords and obtain their search volumes separately in the fixed scaling mode (i.e., the search volumes are scaled to the average traffic during a fixed point in time (usually January 2004)). Then we aggregate the search volumes by assigning their relative weights based on the search frequency of each group with the rank order centroid method [Barron and Barret 1996].

We collected the measurements for network attacks via DShield (www.DShield.org). DShield is a community-based collaborative firewall/intrusion detection log correlation system (see, e.g., Wikipedia.org). DShield receives tens of thousands of firewall and intrusion detection system logs from volunteers throughout the Internet. The logs are reported automatically by client programs running on the submitting hosts (typically once per hour), or anonymously uploaded through the DShield's Web site. Each log entry provided by a network represents one or more packets that violated a local rule. DShield transformed, cataloged, and summarized all of the logs. The aggregated information provides accurate and current snapshots of Internet attacks. It can be used to discover trends in activity, confirm widespread attacks, or assist in preparing better firewall rules. DShield is the data collection engine behind the SANS Internet Storm Center (ISC). It was officially launched in November 2000, and has grown to be a dominating attack correlation engine with worldwide coverage. Analysis provided by DShield has been used in the early detection of several worms, like Ramen, Code Red, Leaves, and SQL Snake. DShield data is also regularly used by researchers to analyze attack patterns.

We derived two measurements of network attacks, intensity and prevalence, based on the DShield reports on Daily Survival Time and those on Daily Data Volume (Submissions/Day) respectively. The daily reports of survival time provide us the daily average time (in the unit of minutes) between probes at a target IP address for five different categories of ports: Windows-specific ports (e.g., File sharing) (denoted as $I_t-WIN, t = 1, \dots, 92$), Unix-specific ports (e.g., dns, ssh) (I_t-UNI), ports which are used (and vulnerable) by applications on various operating systems (I_t-APP), ports used by P2P software (I_t-P2P), and ports that are commonly used by backdoors (I_t-BCK). If we assume that most of these reports are generated by malwares/attacks that attempt to propagate,

Table I. The List of Keywords (regardless of cAsE)

Trojan	Firewall	Phishing	Malware	Adware
Spam	Spyware	patching	authentication	Botnet
identity and theft	access and control	forensic and computer	intrusion and detection	antivirus or (anti and virus)
packet and (filter or filtering)	hack or hacking or hacker	malicious and (code or codes or software)	cracking and (software or computer or network)	denial and of and service
defense and (network or computer)	risk and (computer or information or network or software)	fraud and (software or program or network or internet)	attack and (network or computer or internet or software)	crime and (cyber or computer or internet or electronic or network)
(software or computer or computer)	program or windows or network or internet)	(software or network or internet)	(vulnerability or vulnerabilities)	(software or computer or computer)
(software or computer or program or windows or network or internet or computer)	program or windows or network or internet or computer)	(software or network or internet or computer)	(software or network or internet or computer)	(software or computer or computer)
(software or program or windows or network or internet or computer or cyber)	(software or program or windows or network or internet or computer)	(software or network or internet or computer)	(software or network or internet or computer)	(software or computer or computer)
(software or program or windows or network or internet or computer)	(software or program or windows or network or internet or computer)	(software or network or internet or computer)	(software or network or internet or computer)	(software or computer or computer)
(software or protection) and (software or program or windows or network or internet or computer)	(software or program or windows or network or internet or computer)	(software or network or internet or computer)	(software or network or internet or computer)	(software or computer or computer)



* The data point on May 17, 2006 is to be replaced by the trimmed mean.

Fig. 1. The daily frequency of search for information security knowledge (Y_t).

an unpatched system would be infected by such a probe. Though attacks to different operating systems or applications show different patterns, the survival times for different categories are highly correlated. We performed a Principal Component Analysis (PCA) and kept factors having an Eigenvalue over 1. We used varimax as the method of factor rotation and found that there is only one factor which explains 60.73% of the variance. The loadings of I_t -WIN, I_t -UNI, I_t -APP, I_t -P2P, and I_t -BCK to the PCA factor are .82, .83, .85, .70, and .69 significant at $\alpha = 0.01$. We created the factor score through the regression method (denoted as I_t , $t = 1, \dots, 92$) and used it as a reversed daily measurement of the intensity of network attacks. The measurement provides us a holistic view of network attacks across operating systems and applications.

Our second measurement, the prevalence of network attacks, was derived based on the reports of daily data volume (Submissions/Day). We used a normalized daily number of targets (i.e., daily number of targets divided by daily number of the reports) (denoted as P_t , $t = 1, \dots, 92$) as our measurement of the prevalence of network attacks.

We collected vulnerability disclosures from National Vulnerability Database (NVD) (see <http://nvd.nist.gov/>). NVD is a comprehensive Common Vulnerability Exposure (CVE) vulnerability database that integrates all U.S. Government publicly available vulnerability resources. For each vulnerability, the database has the date when it was initially published and a CVSS⁷ base score ranging from 0 to 10. The score measures vulnerability severity and was calculated based on the inherent characteristics of the vulnerability, such as exploitability

⁷Common Vulnerability Scoring System, <http://www.first.org/cvss/cvss-guide.html>.

Table II. Descriptive Statistics of Variables (92 observations)

Variables	Mean	Std. Dev.	Min	1st Quartile	Median	3rd Quartile	Max
Y_t Daily # of queries	269.86	74.91	2	225	256	312.5	472
H_t Daily # of Vul. in High	5.78	6.06	0	1	4.5	8	29
M_t Daily # of Vul. in Medium	10.65	10.23	0	1.5	8	17	41
L_t Daily # of Vul. in Low	2.05	2.89	0	0	1	3	14
S_t Daily Sum of Vul. Score	12.86	9.22	0	6.65	11.55	18.05	39.00
I_t (PCA factor) Attack Intensity	0.21	1.11	-1.21	-0.35	-0.08	0.52	6.48
I_t -WIN Surv. time for Win ports	39.25	8.96	27	34	38	42	86
I_t -UNI Surv. time for Unix ports	776.34	206.49	468	630	752.5	874	1519
I_t -APP Surv. time for App. Ports	240.39	147.44	97	170.5	202	261.5	1172
I_t -P2P Surv. time for P2P ports	538.35	296.23	200	374.5	458	598	2193
I_t -BCK Surv. time for Backdoor ports	7237.58	4455	2247	4742	6125	8607	38963
P_t ($\times 10^{-3}$) Attack Prevalence	11.14	3.23	1.34	9.63	10.60	12.02	29.88

and impact. In addition to the numeric CVSS scores, NVD provides qualitative severity rankings of “low,” “medium,” and “high” simply mapped from the numeric CVSS scores: vulnerabilities are labeled “low” severity if they have a CVSS base score of 0.0–3.9, “medium” severity if 4.0–6.9, and “high” severity if 7.0–10.0. We derived two alternative measures for vulnerability disclosures. We first counted the daily number of vulnerabilities in CVSS severity high (denoted as H_t), medium (M_t), low (L_t) respectively in order to examine the possible different impacts they might have. We also summarize the numeric CVSS scores of all vulnerabilities announced in a day (S_t) as an alternative measure for the severity of vulnerability announced in a day.

Table II presents the descriptive statistics of these variables. We noticed that the 78th observation of Y_t (May 17, 2006, with a value of 2) stands out for the rest of the data based on its histogram. It is remarkably lower than both earlier and later observations. Similar occurs for the 55th observation of P_t (April 24, 2006, with a value of 1.34×10^{-3}). The 91th and 92th observations (May 30 and May 31, 2006) of I_t (with values of 27.55×10^{-3} and 29.88×10^{-3}) and P_t (with values of 6.48 and 5.63) are extremely high compared with the rest of the corresponding series. As we do not have reasonable explanations on such extremal variations, we replaced these extremes by the trimmed mean, which is the mean of the whole series excluding those extremes [Anscombe 1960].

3.2 Empirical Modeling

We can see that the frequency of the search for information security knowledge varies significantly from day to day in Figure 1. The idea here is that changes in H_t , M_t , L_t , (or alternatively S_t), I_t , and P_t may signal current or later changes in Y_t . In other words, maybe movements in Y_t (daily number of queries) reflect movements in H_t , M_t , L_t , (or alternatively S_t), I_t , and P_t . In this section, we introduce our modeling approach and estimation results.

3.2.1 *A Dynamic Regression Modeling Approach.* As both dependent and independent variables follow a time series, we use dynamic regression models [Pankratz 1991] to model the response of search behavior to vulnerability disclosures and network attacks. The general notation of the regression functions for these models can be written as

$$Y_t = C + \sum_{j=1}^J \sum_{k=0}^{L_j} \beta_{jk} X_{j(t-k)} + N_t, \quad (1)$$

where Y is the dependent variable, X represents a set of independent variables that could include (H, M, L, S, I, P) , t is the time when the dependent variable was observed, J is the total number of independent variables we include, L_j is the total number of time lags we include for the j th independent variable, β_{jk} is the regression coefficient, C is a constant term, and N_t is the disturbance of the regression function involving time series data. Note that N_t could be autocorrelated and its behavior might be described by an ARIMA (p, d, q) process which has p autoregressive order, d differencing order, and q moving average terms.

We include the smallest L_j ($L_j \geq 0$) for j th independent variable that has a significant impact on the dependent variables in our final models. In order to estimate the underlying ARIMA process for N_t , we used unit-root tests [Dickey and Fuller 1979, 1981] to examine whether the residuals of a regression, $\hat{N}_t = Y_t - \hat{C} - \sum_{j=1}^J \sum_{k=0}^{L_j} \hat{\beta}_{jk} X_{j(t-k)}$, is stationary (i.e., whether the means and variance of the process are time independent). If \hat{N}_t is not stationary, we take differences on both output and inputs in Eq. (1), and then reestimate \hat{N}_t . We repeat the procedure until the residuals are stationary. We relied on correlograms (including autocorrelations and partial correlations) and Q-statistics to detect autoregressive order and the number of moving average terms. Please refer to Box et al. [1994] for a detailed discussion on various patterns of correlograms. We also use Breusch-Godfrey Serial Correlation LM test to confirm serial correlations existing in the residual. We follow Pankratz [1991] for our overall modeling strategy. We used Akaike Information Criteria (AIC) and Schwarz Criterion (BIC) as the metrics for model fitting and variables selection.

3.2.2 *Estimation Results.* We first checked feedback from the dependent variable Y_t to independent variables I_t and P_t . In other words, we checked whether more search on information security will lead to more network attacks, as the Internet could be a place for malicious users to gain attacking knowledge and tools, and initiate, for example, copycat attacks. (We may reasonably assume that search behavior may not affect vulnerability disclosures). As we argued before, the primary cause of search for information security knowledge is the users' protection and defense intention. However, two other arguments that deal with the other side of the coin can be put forth. First, there is a possibility that users of search engines who search for information security knowledge are actually hackers [Moore and Clayton 2009]. For instance, if we assume that there are a fair number of hackers, even though most hackers

Table III. Regression Coefficients to Estimate Feedback from Y_t to I_t and P_t

K	0	1	2	3
$\hat{b}_{3k} (*10^{-4})$	4.96	5.11	5.13	-6.04
$ t $	0.76	0.76	0.77	0.93
$\hat{c}_{3k}(*10^{-4})$	5.00	-1.01	-10.97	37.77
$ t $	0.27	0.05	0.58	2.07

have knowledge of a certain level of information security technology [Chantler 1995], it may still be necessary for them to perform a search for information security knowledge via a search engine to mount further attacks. Second, as Karnow [2007] points out there is a growing interest in strike-back tools that automate counterstrikes to Internet security attacks. So it is possible to reason that those users who are searching for security information are actually victims of such attacks who may wish to apply the doctrine of self-defense to carry out a counterstrike. Our feedback check will help us to avoid the inappropriate use of a single-equation model as specified in Eq. (1) if there is feedback. We estimate the following two equations.

$$I_t = C + \sum_{k=1}^3 b_{1k} I_{t-k} + \sum_{k=0}^3 b_{2k} P_{t-k} + \sum_{k=0}^3 b_{3k} Y_{t-k} + a_t \quad (2)$$

$$P_t = C + \sum_{k=0}^3 c_{1k} I_{t-k} + \sum_{k=1}^3 c_{2k} P_{t-k} + \sum_{k=0}^3 c_{3k} Y_{t-k} + a_t \quad (3)$$

We are especially interested in the \hat{b}_{3k} and \hat{c}_{3k} coefficients since these measure possible feedback from Y_t to I_t and P_t , respectively. These coefficients and their t -values are shown in Table III. Only \hat{c}_{33} is significant. This one significant coefficient can easily be attributed to sampling variation. The F statistics were calculated for both Eq. (2) and Eq. (3) to test the null hypotheses that $b_{30} = b_{31} = b_{32} = b_{33} = 0$ and $c_{30} = c_{31} = c_{32} = c_{33} = 0$. We have $F = 0.76$ for Eq. (2) and $F = 1.35$ for Eq. (3). Both are insignificant at the level of 5%. It is reasonable to conclude that there is no feedback from the output to the inputs. This probably also implies that hackers do not use search engines to search for information that will help them in hacks. Instead they may go directly to certain Web sites or online forums dedicated to hackers to share their experiences, exchange attacking tools, or discuss attacking techniques [Armstrong and Forde 2003]. In other words, instead of conducting a search for information security knowledge via search engines, they may directly visit specific Web sites or join hacker communities to obtain the information they need. It probably also implies that counterstrikes are not yet sufficiently large to make an impact on the quantum of attacks. We therefore proceed to build a single-equation model.

Table IV summarizes the estimation results for six alternative models that we had. In Model I, H_t , M_t , L_t , I_t , $I_{(t-1)}$, P_t , and $P_{(t-1)}$ were included in the regression. An ARIMA (2, 0, 0) process was identified for the regression residual

Table IV. Estimation Results

Variables	Model I	Model II	Model III	Model IV	Model V	Model VI
H_t	-0.16 (1.80)	-	0.78 (0.93)	-	-	-
M_t	1.17 (1.01)	-	-	0.68 (0.55)	-	-
L_t	-2.23 (3.32)	-	-	-	0.78 (2.16)	-
S_t	-	0.65 (0.62)	-	-	-	-
I_t	-45.42** (16.32)	-44.14** (16.40)	-45.00** (16.36)	-45.69** (16.28)	-44.52** (16.46)	-44.06** (16.27)
$I_{(t-1)}$	-37.26* (16.69)	-43.45** (16.55)	-42.49** (16.44)	-39.84* (16.49)	-43.75** (16.43)	-43.99** (16.31)
P_t ($\times 10^{-3}$)	13.98** (5.08)	14.11** (5.15)	13.83** (5.11)	13.95** (5.07)	14.05** (5.16)	14.09** (5.13)
$P_{(t-1)}$ ($\times 10^{-3}$)	11.03* (5.10)	11.57* (5.23)	11.52* (5.11)	11.09* (5.09)	11.55* (5.17)	11.66* (5.12)
$AR(1)$	0.24* (0.11)	0.22* (0.11)	0.23* (0.11)	0.23* (0.11)	0.22* (0.11)	0.21* (0.11)
$AR(2)$	0.29** (0.11)	0.27** (0.11)	0.25* (0.11)	0.27** (0.11)	0.27* (0.11)	0.27** (0.11)
Model Assessment						
AIC	11.00	10.99	10.97	10.97	10.98	10.96
BIC	11.26	11.18	11.17	11.16	11.18	11.13

- excluding the variable from the regression.

** coefficient significant at 0.01.

* coefficient significant at 0.05.

(•): standard error of the coefficient.

(which is the same for the other models). While we noticed that the daily number of vulnerability disclosures in different severity levels cannot significantly affect search behavior, we used the alternative measure, S_t , for vulnerability disclosures in our Model II. We did not find the significant effect either. We then included only one of H_t , M_t , L_t sequentially in our regressions (Model III, Model IV, Model V). None of them was found to have a significant impact on search behavior. We finally excluded the variables that measure vulnerability disclosures in the regression (Model VI), and found that AIC and BIC are not worse. For robustness checking, we run dynamic regression models by only including vulnerabilities announcement H_t , M_t , L_t (or S_t) with time lags from 0–3 and from 0–7 as independent variables. In both cases we found no term significant. Thus we conclude that vulnerability disclosures do not significantly impact individuals' search frequency for information security-related knowledge, supporting H3(a2) and H3(b2).

Across all these six models, we found that both attack intensity (I) and attack prevalence (P) significantly impacts the frequency of search for information security knowledge. The more intense network attacks, the more frequent the search, supporting H1(a). And the more prevalent network attacks, the more frequent the search, supporting H1(b). Network attacks on the current day and

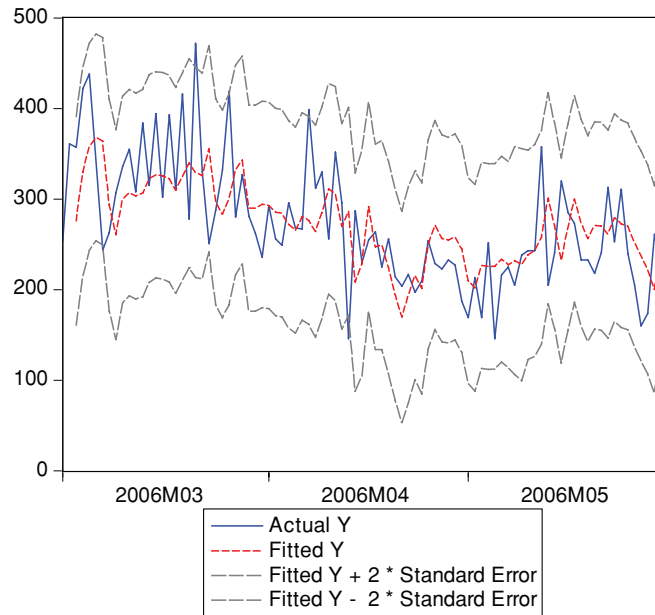


Fig. 2. Actual and fitted Y with two standard error bounds (Model VI).

one day prior significantly impact the search frequency for information security. In other words, network attacks have a short-term but immediate impact on search frequency. Thus H2(a) and H2(b) are supported.

Figure 2 plots the actual series and the fitted series of Y with ± 2 standard error bounds based on the estimation results of Model VI. Almost all of actual Y s fall in the ± 2 standard error bounds. For the purpose of robustness checking, we also used a normalized daily frequency of search for information security knowledge, which is the number of queries for information security divided by the number of all queries occurring in the same day, as our dependent variables in our regressions. We found similar results as presented earlier.

3.2.3 An Event Study of the Impact of Outbreaks of Notable Viruses With Google Trends Data. To complement our analysis, we further conducted an event study to show how the outbreaks of notable viruses impact information security-related search with the data from Google Trends. Event study has been used in accounting, finance, and economics to study the impact of a specific event or announcement on the value of a firm. For an introduction of event study analysis please refer to Campbell et al. [1996] and McWilliams and Siegel [1997].

As event study analysis is suitable to understand how rare events affect individual behaviors, in our analysis we define the events as the outbreaks of notable computer viruses and explore how such outbreaks will impact the search volume of related keywords in Google. A variety of computer viruses were identified by antivirus vendors almost every day, and most of them only have limited effect or low severity, the outbreaks of which cannot be considered

as rare events. We focused on the viruses having affected a significant amount of ordinary users. We identified the name of these viruses using the virus list provided by Wikipedia which includes 23 of the most notable viruses/worms.⁸ We collected the dates on which the viruses were discovered from antivirus vendors including Symantec and McAfee. Among these viruses, we removed the ones which were discovered before April 1, 2004 to assure that we have search data from Google Trends to calculate the abnormal changes in information security-related search. To avoid compounding effects, we also removed the viruses which were discovered within 20 days of another notable virus. In addition, we removed the viruses whose discovery dates were not consistent among different antivirus vendors. Finally, 11 viruses/worms were used in our event study.

We collected the daily search volume that occurred in the USA from Google Trends (www.google.com/trends) related to virus/worms with the following search string:

```
'Trojan|(network worms)|(internet worms)|(computer worms)|(malicious code)
|(malicious codes)|(software virus)|(computer virus)|(program virus)|(windows
virus)|(internet virus)|(network virus)|malware|antivirus|(anti virus)| (com-
puter protection)|patching'
```

With the preceding string, Google Trends gave us the aggregate daily search volumes after January 1, 2004 of the queries separated by “|”.

Analogous to the return of stock price in traditional event study, we used “search volume changes” to capture the trend of search volume. The search volume change on day t is denoted as R_t and is calculated as $R_t = (D_t - D_{t-1})/D_{t-1}$, where D_t is the volume of the search queries on day t .

Two models have been used in traditional event study literature to estimate the abnormal return on day t : market adjusted model and mean adjusted model. In the market adjusted model, the expected return of stock m on day i is estimated using the portfolio of stocks such as Standard & Poor’s 500 or a market index. As there is no such portfolio existing in the search volume data, we used the mean adjusted model to estimate the abnormal search volume change. Following Brown and Warner [1985], the expected search volume change of virus i on day t is estimated by taking the average of the “search volume changes” 109 to 10 days prior to the day that the virus (estimation period) was discovered

$$\bar{R}_i = \frac{1}{100} \sum_{t=-109}^{t=-10} R_{it},$$

where R_{it} is the search volume change of virus i on day t , and \bar{R}_i is the expected search volume changes of virus i over the estimation period. The abnormal search volume changes on day t (AR_{it}) is calculated as

$$AR_{it} = R_{it} - \bar{R}_i.$$

⁸see http://en.wikipedia.org/wiki/Timeline_of_computer_viruses_and_worms for details.

Table V. Abnormal Search Volume Changes

Day	-4	-3	-2	-1	0	1	2	3	4
AR (t-statistics)	0.03 (0.41)	-0.00 (-0.03)	-0.11 (-1.76)	-0.00 (-0.01)	-0.01 (-0.23)	0.12* (2.06)	-0.01 (-0.21)	0.02 (0.35)	-0.01 (-0.17)

*Significant at 0.05.

The t-statistics for any event day t is calculated as $t - statistics = \frac{\overline{AR}_t}{S(\overline{AR}_t)}$, where $\overline{AR}_t = \frac{1}{N} \sum_{i=1}^N AR_{it}$, N is the number of viruses, and $S(\overline{AR}_t)$ is the standard deviation of \overline{AR}_t . $S(\overline{AR}_t)$ is estimated as $\widehat{S}(\overline{AR}_t) = \sqrt{(\sum_{t=-109}^{t=-10} (\overline{AR}_t - \overline{\overline{AR}})^2) / 100}$, where $\overline{\overline{AR}} = \frac{1}{100} \sum_{t=-109}^{t=-10} \overline{AR}_t$.

The results of the event study are summarized in Table V. We calculated the abnormal search volume changes of days surrounding the outbreaks of the viruses, and their corresponding t statistics. Our results show that the abnormal search volume changes of day 1 (one day after the outbreaks of viruses) is positive and significant at the level of 0.05, and the abnormal search volume changes of day 2 is not significant. This result indicates that the search for information security-related topics increases on the first day following when the viruses are discovered, and the impact goes down after that. This result is consistent with our finding in Section 3.2.2.

4. CONCLUSION

In this study, we investigate the drivers of search for information security knowledge and explore how network attacks and vulnerability disclosures affect users' search for information security knowledge. Our findings suggest that both intensity and prevalence of network attacks are positively related to the frequency of search for information security knowledge. Network attacks on the current day and one day prior drive search behavior. Users actively acquire information security-related knowledge when they are experiencing attacks. This finding is consistent with healthcare literature which finds that the spread of a biological virus is correlated with the frequency of patients' search. This result reveals that people present similar behavior patterns in both physical and cyber worlds when they are facing threats.

We did not find a significant relationship between vulnerability disclosures and search for information security knowledge. While vulnerability disclosures may accelerate patch release and make information security professionals take proactive protection actions, they do not lead ordinary users to seek more information security knowledge. The limited reactions of ordinary users to vulnerability disclosures can be due to different reasons, such as dissemination channel and the technical nature of the announcements. Further users may not perceive vulnerabilities as immediate threats or as network attacks, and be less motivated to search [Liang and Xue 2009].

4.1 Implications for Theory

First, our study contributes to vulnerability disclosure literature. While prior studies focus on disclosure policy of the public agencies like CERT, disclosure

mechanisms, vulnerability market [Radianti et al. 2009a, 2009b] and how attackers and vendors respond to the disclosures [Arora et al. 2010; Cavusoglu et al. 2007], we study the impacts of vulnerability disclosures on the search behavior of ordinary users for information security. We find that vulnerability disclosures draw less attention from ordinary users.

Second, while most research on information security rely on user self-reported data, our study uses secondary data which consists of search engine logs, firewall and intrusion detection systems logs and vulnerability announcements. This dataset provides us the opportunity to study users' actual information security behaviors rather than their intentions. The dataset provides us more objective measurements on user behaviors, and allows us to study public reaction to information threats instead of a specific group.

Third, threat avoidance theory [Liang and Xue 2009] provides a new perspective to investigate users' information security behavior. However, few studies have empirically tested this model. Our research provides empirical evidence to support part of the threat avoidance theory model by demonstrating that the intensity of malicious attacks could determine users' threat perception, which in turn influences users' avoidance behaviors. On the other hand, vulnerability disclosures do not cause immediate damages and may not be perceived as severe, immediate threats like ongoing attacks; users are less motivated to take actions.

Our results reveal that search for information security changes over time. In addition to temporal dimension, future research could investigate information security-related search along the geographical dimensions. Recently, search queries have been used to monitor flu outbreaks across different regions (see <http://www.google.org/flutrends/> for details). It would be interesting to investigate whether search for information security follows a similar pattern and changes as network attacks spread over the Internet.

4.2 Implications for Practice

First, our study has implications for information security awareness campaign and knowledge dissemination. When prompting information security knowledge among ordinary users, it is important to help users realize the threats they are or will be confronting. Recognizing the severity of threats will provide users a better motivation for information security knowledge. And they may more proactively engage in knowledge seeking for information security.

As the public usually ignores warnings on information security vulnerabilities and fails to take proactive actions to protect their systems, our study suggests that in order to persuade users to take actions to improve their systems security, the messages delivered to them need to show a high possibility of being affected and the associated serious consequences. Similarly, to attract users to purchase the antivirus software packages, software vendors not only need to show the effectiveness of their products in defense of systems, but also how real and immediate the threat is to users.

Second, our results also provide implications for software vendors and security countermeasures vendors. It is important for software vendors and

security measurement vendors to realize the fact that the users react to ongoing attacks, but not vulnerability disclosures. In other words, when users are motivated to take proper actions to fix their software vulnerabilities and update their detection signatures of their countermeasures it may be too late. Thus, when the vendors design their software and countermeasures, they could build in proper automatic mechanisms (such as autoupdated functionalities) without much user intervention to fix software vulnerabilities, distribute patches, or update detection signatures of countermeasures. Furthermore, if the vulnerability is really severe, the vendors may directly warn their customers of its consequences and provide detailed instructions to fix it.

Third, our study suggests the search queries on information security submitted to the search engine could be used to monitor and discover network attack epidemics as an alternative method. Currently, most network attack monitoring systems use the logs of antivirus/firewall software installed in users' systems to keep track of network attack epidemics. Using search queries provides us an alternative way to track the epidemics of network attacks. It might provide us a more holistic view of the attack activities over the network across platforms and operating systems.

Fourth, our findings show that network attacks trigger computer users' search for information security-related knowledge. On the one hand, this finding indicates that users have a high information need subsequent to network attacks. On the other hand, it also implies users do not have reliable sources to gain necessary knowledge about information security and have to use general search engines to seek for it. Although search engines have lots of advantages in locating information, their results sometimes rely on the quality of the queries entered and the experiences or knowledge of the searchers. For example, experienced users may take less time to find information than novice users. In contrast, Web portals on information security knowledge could provide users with the related knowledge in a convenient and systematic way, especially for ordinary users. Therefore, it will be beneficial if organizations and the government build vertical portals specific to distribute information security-related knowledge. Currently, although such kinds of portals exist, the information on these portals is either limited or not updated frequently to reflect new information security threats. Increasing the number and the quality of information security-related portals and Web sites could improve the effectiveness and efficiency of protecting information security.

Finally, this article has several limitations that can be studied in future research. First, while the use of the actual search logs is a strength, the search logs cover a very limited time period (three months in 2006). The question still remains: Can we assume that this short window is representative of search behavior in general? In addition, in terms of the sample, using only AOL client software searches is a limitation, since we cannot be certain that AOL searches are representative of all search behavior (though Google was the AOL search engine at the time of the data collection). It would be useful to compare AOL data versus other search engines. This is necessary for generalizability purposes. Second, the article used the dynamic regression approach where all search terms were aggregated into one single variable/model. We also used an

event analysis study on the data from Google Trends to understand how the outbreaks of notable viruses will affect the search for virus-related information. It would be useful to repeat the analysis for some other specific and highly impactful network attacks (and the search frequencies of words pertaining to those attacks only) and observe the differences between the aggregate and the specific. Third, in this article there is an assumption of static user response to stimuli of events. Others have argued that security behavior is itself adaptive, so that users might become adjusted to higher levels of security risk over time. Such behaviors take much longer than three months to establish, and would be masked in this dataset. Finally, while the topic of this research is ordinary users, we cannot be certain that the search logs contain only searches by ordinary users, and not IS security people or other technical users. In other words, it is not possible to partial out any search logs that do not belong to ordinary users.

ACKNOWLEDGMENTS

The authors thank the SE, AE and referees for comments that have greatly improved the article.

REFERENCES

- ANDERSON, C. L. AND AGARWAL, R. 2006. Practicing safe computing: Message framing, self-view, and home computer user security behavior intentions. In *Proceedings of the 27th International Conference on Information Systems*.
- ANDERSON, C. L. AND AGARWAL, R. 2010. Practicing safe computing: A multimethod empirical examination of home computer user security behavioral Intentions. *MIS Quart.* 34, 3, 613–643.
- ANSCOMBE, F. J. 1960. Rejection of outliers. *Technomet.* 2, 2, 123–147.
- ARMSTRONG, H. L. AND FORDE, P. J. 2003. Internet anonymity practices in computer crime. *Inform. Manag. Comput. Secur.* 11, 5, 209–215.
- ARORA, A., KRISHNAN, R., TELANG, R., AND YANG, Y. 2010. An empirical analysis of software vendors' patch release behavior: Impact of vulnerability disclosure. *Inf. Syst. Res.* 21, 1, 115–132.
- ARORA, A., NANDKUMAR, A., AND TELANG, R. 2006. Does information security attack frequency increase with vulnerability disclosure? An empirical analysis. *Inf. Syst. Frontiers* 8, 5, 350–362.
- BARRON, F. H. AND BARRETT, B. E. 1996. Decision quality using ranked attribute weights. *Manag. Sci.* 42, 11, 1515–1523.
- BATTELLE, J. 2005. *The Search*. Penguin Books, New York.
- BOX, G. E. P., JENKINS, G. M., AND REINSEL, G. C. 1994. *Time Series Analysis: Forecasting and Control*, Prentice Hall, Englewood Cliffs, NJ.
- BROUWERS, M. C. AND SORRENTINO, R. M. 1993. Uncertainty orientation and protection motivation theory: The role of individual differences in health compliance. *J. Person. Social Psych.* 65, 1, 102–112.
- BROWN, S. J. AND WARNER, J. B. 1985. Using daily stock returns the case of event studies. *J. Finan. Econ.* 14, 3–31.
- BROWNE, G. J., PITTS, M. G., AND WETHERBE, J. C. 2007. Cognitive stopping rules for terminating information search in online tasks. *MIS Quart.* 31, 1, 89–104.
- CAMPBELL, J. Y., LO, A. W., AND MACKINLAY, A. C. 1996. *The Econometrics of Financial Markets*, Princeton University Press, Princeton, NJ.
- CAVUSOGLU, H., CAVUSOGLU, H., AND RAGHUNATHAN, S. 2007. Efficiency of vulnerability disclosure mechanisms to disseminate vulnerability knowledge. *IEEE Trans. Softw. Engin.* 33, 3, 171–185.
- CHANTLER, A. N. 1995. Risk: The profile of a computer hacker. Tech. rep., Curtin University, Perth.

- CHAU, M., FANG, X., AND SHENG, O.R.L. 2007. What are people searching on government web sites? *Comm. ACM* 50, 4, 87–92
- CHOU, P. H. B. AND WISTER, A. V. 2005. From cues to action: Information seeking and exercise self-care among older adults managing chronic illness. *Canad. J. Aging* 24, 4, 395–408.
- COOPER, C. P., MALLON, K. P., LEADBETTER, S., POLLACK, L. A., AND PEIPINS, L. A. 2005. Cancer Internet search activity on a major search engine, United States 2001–2003. *J. Med. Internet Res.* 7, 3, e6.
- DERVIN, B. 1983. An overview of sense-making research: Concepts, methods and results to date. In *Proceedings of the International Communications Association Annual Meeting*.
- DICKEY, D. AND FULLER, W. 1979. Distribution of the estimators for autoregressive time series with a unit root. *J. Amer. Statist. Assoc.* 74, 427–431.
- DICKEY, D. AND FULLER, W. 1981. Likelihood ratio tests for autoregressive time series with a unit root. *Econometr.* 49, 4, 1057–1072.
- DOWLAND, P. S., FURNELL, S. M., ILLINGWORTH, H. M., AND REYNOLDS, P. L. 1999. Computer crime and abuse: A survey of public attitudes and awareness. *Comput. Secur.* 18, 8, 715–729.
- EPPIRIGHT, D. R., JOHN F. TANNER, J., AND HUNT, J. B. 1994. Knowledge and the ordered protection motivation model: Tools for preventing AIDS. *J. Bus. Res.* 30, 1, 13–24.
- ETTREDGE, M., GERDES, J., AND KARUGA, G. 2005. Using web-based search data to predict macroeconomic statistics. *Comm. ACM* 48, 11, 87–92
- FLOYD, D. L., PRENTICE-DUNN, S., AND ROGERS, R. W. 2000. A meta-analysis of research on protection motivation theory. *J. Appl. Soc. Psych.* 30, 2, 407–429.
- FURNELL, S. M., JUSOH, A., AND KATSABAS, D. 2006. The challenges of understanding and using security: A survey of end-users. *Comput. Secur.* 25, 1, 27–35.
- GINSBERG, J., MOHEBBI, M. H., PATEL, R.S., BRAMMER, L., SMOLINSKI, M.S., AND BRILLIANT, L. 2008. Detecting influenza epidemics using search engine query data. *Nature* 457, 19, 1012–1014.
- GRIFFIN, R. J., DUNWOODY, S., AND NEUWIRTH, K. 1999. Proposed model of the relationship of risk information seeking and processing to the development of preventive behaviors. *Environ. Res.* 80, 2, S230–S245.
- HERATH, T. AND RAO, H. R. 2009a. Encouraging information security behaviors in organizations: Role of penalties, pressures and perceived effectiveness. *Decis. Support Syst.* 47, 2, 154–165.
- HERATH, T. AND RAO, H. R. 2009b. Protection motivation and deterrence: A framework for security policy compliance in organisations. *Euro. J. Inf. Syst.* 18, 2, 106–125.
- HUSSAIN, A., HEIDEMANN, J., AND PAPADOPOULOS, C. 2003. A framework for classifying denial of service attacks. In *Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*. 99–110.
- JANSEN, B. J. 2007. The comparative effectiveness of sponsored and nonsponsored links for web E-commerce queries. *ACM Trans. Web* 1, 1, 3-es.
- KARNOW, C. E. A. 2007. Counterstrike. In *Cybercrime: Digital Cops in A Networked Environment*, J.M. Balkin, K. Eddan, J. Grimmelmann, N. Kozlovski, S. Wagman and T. Zarsky, Eds. NYU Press, New York.
- KEPHART, J. O. AND WHITE, S. R. 1993. Measuring and modeling computer virus prevalence. In *Proceedings of the IEEE Symposium on Security and Privacy*. 2–14.
- KUMAR, N., MOHAN, K., AND HOLOWCZAK, R. 2008. Locking the door but leaving the computer vulnerable: Factors inhibiting home users' adoption of software firewalls. *Decis. Support Syst.* 46, 1, 254–264.
- LI, P. AND RAO, H. R. 2007. An examination of private intermediaries' roles in software vulnerabilities disclosure. *Inf. Syst. Frontiers* 9, 5, 531–539.
- LIANG, H., AND XUE, Y. 2009. Avoidance of information technology threats: A theoretical perspective. *MIS Quart.* 33, 1, 71–90.
- MANIKOPOULOS, C. AND PAPAVALASSIOU, S. 2002. Network intrusion and fault detection: A statistical anomaly approach. *IEEE Comm. Mag.* 40, 10, 76–82.
- MCQUEEN, M. A., MCQUEEN, T. A., BOYER, W. F., AND CHAFFIN, M. R. 2009. Empirical estimates and observations of 0day vulnerabilities. In *Proceedings of the 42nd Hawaii International Conference on System Sciences*.
- MCWILLIAMS, A. AND SIEGEL, D. 1997. Event studies in management research: Theoretical and empirical issues *Acad. Manag. J.* 40, 3, 626–657.

- MELL, P., SCARFONE, K., AND ROMANOSKY, S. 2007. *A Complete Guide to The Common Vulnerability Scoring System Version 2.0*, Forum of Incident Response and Security Teams.
- MILLER, G. A. 1983. Informavores. In *The Study of Information: Interdisciplinary Messages*, F. Machlup and U. Mansfield, Eds. Wiley, New York, 111–113.
- MOORE, T. AND CLAYTON, R. 2009. Evil searching: Compromise and recompromise of Internet hosts for phishing. In *Proceedings of the 13th International Conference on Financial Cryptography and Data Security*.
- NEUWIRTH, K., DUNWOODY, S., AND GRIFFIN, R. J. 2000. Protection motivation and risk communication. *Risk Anal.* 20, 5, 721–734.
- NG, B.-Y., KANKANHALLI, A., AND XU, Y. C. 2009. Studying users' computer security behavior: A health belief perspective. *Decis. Support Syst.* 46, 4, 815–825.
- PANKRATZ, A. 1991. *Forecasting with Dynamic Regression Models*. John Wiley & Sons, New York.
- PASS, G., CHOWDHURY, A., AND TORGESON, C. 2006. A picture of search. In *Proceedings of the 1st International Conference on Scalable Information Systems (InfoScale '06)*.
- PNG, I., WANG, C., AND WANG, Q. 2008. The deterrent and displacement effects of information security enforcement: International evidence. *J. Manag. Inf. Syst.* 25, 2, 125–144.
- RADIANTI, J., GONZALEZ, J. J., AND RICH, E. 2009a. A quest for a framework to improve software security: Vulnerability black markets scenario. In *Proceedings of the 27th International Conference of the System Dynamics Society*.
- RADIANTI, J., RICH, E., AND GONZALEZ, J. J. 2009b. Vulnerability black markets: Empirical evidence and scenario simulation. In *Proceedings of the 42nd Hawaii International Conference on System Sciences*.
- RICHARDSON, M. 2008a. Learning about the world through long-term query logs. *ACM Trans. Web* 2, 4, Article 21.
- RICHARDSON, R. 2008b. *CSI Computer Crime & Security Survey*. Computer Security Institute.
- RIPPETOE, P. A., AND ROGERS, R. W. 1987. Effects of components of protection-motivation theory on adaptive and maladaptive coping with a health threat. *J. Person. Social Psych.* 52, 3, 596–604.
- ROGERS, G. O. 1997. The dynamics of risk perception: How does perceived risk respond to risk events? *Risk Anal.* 17, 6, 745–757.
- ROGERS, R. W. 1975. A protection motivation theory of fear appeals and attitude change. *J. Psych.* 91, 1, 93–114.
- ROSE, D. E. AND LEVINSON, D. 2004. Understanding user goals in web search. In *Proceedings of the International World Wide Web Conference*.
- SEVERTSON, D. J., BAUMANN, L. C., AND BROWN, R.L. 2006. Applying a health behavior theory to explore the influence of information and experience on arsenic risk representations, policy beliefs, and protective behavior. *Risk Anal.* 26, 2, 353–368.
- SMITH, V. K., DESVOUSGES, W. H., AND PAYNE, J. W. 1995. Do risk information programs promote mitigating behavior? *J. Risk Uncert.* 10, 3, 203–221.
- SRINIVASAN, N. AND RATCHFORD, B. T. 1991. An empirical test of a model of external search for automobiles. *J. Consum. Res.* 18, 233–242.
- TANCER, B. 2008. *Click: What Millions of People Are Doing Online and Why It Matters*. Hyperion, New York.
- VENKATSUBRAMANYAN, S. AND KWAN, S. K. 2008. A web search model for strategic decision making. *AIMS Int. J. Manag.* 2, 3, 197–214.
- WOON, I. M. Y., TAN, G. W. AND LOW, R. T. 2005. A protection motivation theory approach to home wireless security. In *Proceedings of the 25th International Conference on Information Systems*.
- WORKMAN, M., BOMMER, W. H., AND STRAUB, D. 2008. Security lapses and the omission of information security measures: A threat control model and empirical test. *Comput. Hum. Behav.* 24, 6, 2799–2816.
- YBARRA, M. L. AND SUMANB, M. 2006. Help seeking behavior and the Internet: A national survey. *Int. J. Med. Inf.* 75, 1, 29–41.

Received October 2009; revised September 2010; accepted October 2010