INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

THE MATURITY OF CYBERCRIME MARKETS: A SURVEY

PROF. P. B. NIRANJANE, NILESH V. WANKHEDE

1. Assistant Professor, Department of CSE, B.N.C.O.E., Pusad, India.
2. M.E. Student, Department of CSE, B.N.C.O.E., Pusad, India.

Accepted Date: 03/12/2015; Published Date: 01/01/2016

Abstract: - Cybercrime is notoriously maintained and empowered by the underground economy, manifested in black markets. In such markets, attack tools and vulnerability exploits are constantly traded. They should make the effects that we observe every day impossible to sustain. In this paper we show that the market structure and design used by cyber criminals have evolved towards a market design that is similar to legitimate, thriving, in-line forum markets such as eBay. Accordingly, it presents unprecedented opportunities for researchers to tap into these underground cybercriminal communities to develop better insights about collaborative cybercrime activities so as to combat the ever increasing number of cybercrimes.

Keywords: Cybercrime, black markets, security economics, vulnerabilities, exploits, Evilseed, HIC.

Corresponding Author: PROF. P. B. NIRANJANE

Access Online On:

www.ijpret.com

How to Cite This Article:

P. B. Niranjane, IJPRET, 2015; Volume 4 (5): 28-40

PAPER-QR CODE

INTRODUCTION

The burgeoning rise in illicit cyber activity over the last several decades has drawn significant attention from law enforcement, corporate IT departments, and the public-at-large. The common reaction, which springs from a natural desire to protect and defend people, property, and society, is to implement increasingly sophisticated technology barriers. All for the purpose of keeping the high-tech criminal at bay. While this pragmatic outcome-focused reaction is understandable and necessary given the serious stakes involved, its long term strategic efficacy is debatable, considering that cybercrime activity and impacts continue unabated.

This paper takes a different tack, by exploring root causes of deviant cyber-behavior. First, what is cybercrime and how does it differ from "traditional" crime? Second, what draws people to engage in cybercrime in the first place? Are the underlying motives of high-tech criminals identical to those of traditional offenders, or are any unique motives at play? Understanding what triggers individual to engage in cybercriminal activity has important theoretical implications, especially in the fields of cyber-criminology and information systems security.

TECHNOLOGY TRENDS:

Two more cyber game changers will result from technology trends already occurring but predicted to grow exponentially.

A] Big data analytics:

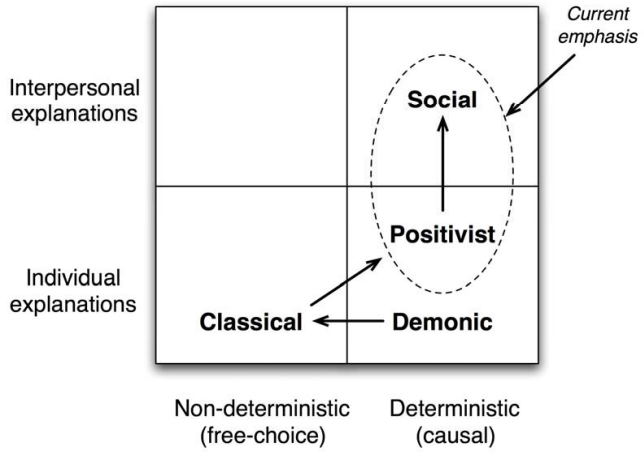
Though still immature from a cyber security perspective, big data analytics (predictive and autonomous) is an area already exerting a noticeable influence. Potentially reaching global scale, able to anticipate multiple new cyber threats within actionable timeframes, and requiring little or no human cyber analysis, big data analytics is a third game changer that could bring new potency to cyber defense.

B] Resilient self-adaption:

Potential innovation based on resilient self-adaptation represents a fourth game changer. Cyber security in this case will derive largely from system agility, moving-target defenses, cyber maneuvering and other autonomous or semi-autonomous behaviors. Exploiting such self-adaptation might mean shifting a significant fraction of design resources from reducing vulnerabilities to increasing resiliency.

CRIMINAL MOTIVES:

The broad philosophical perspectives on criminal motives described above specify individual versus interpersonal, and determinist versus non-determinist, foci. These are summarized in Figure 1, with the arrows approximating the evolution of thought.



When examining cybercrime motives, it is important [4] to note that individuals who possess advanced skills and open access to sophisticated information and communication technologies (ICT) have historically possessed relatively higher levels of education and income. Academic research has also shown strong negative links between criminal activity and socioeconomic indicators such as education and income The foregoing argument invites a nondeterministic approach to the study of cybercrime in effect, sliding the social and positivist perspectives in Figure to the left, to explore alternative nondeterministic explanations (whether individual or interpersonal). In adopting this position, the positivist or social explanations are not being dismissed. The arguments that individuals are influenced to engage in deviant behavior due to preexisting individual factors (e.g., genetics, personality traits) or social factors (e.g., socioeconomic inequities, observational modeling) are quite plausible.

HYPOTHESES ON FORUM MARKETS:

Both Carders.de and HackMarket.ru are forum-based markets. They have administrators, moderators, users' registration procedures, reputation mechanisms and so on. The major difference with Alibaba, eBay, or Craiglist is that they mostly advertise 'illegal' goods. Carders.de specialized mostly in credit cards, while HackMarket.ru specialized mostly in cyber-crime tools, albeit some transactions were also about monetary goods (e.g. credentials for Skype accounts).

To create 'safe trading places' where only experienced and trustworthy users participate, forum-based markets have created a number of mechanisms aimed at distinguishing 'good' and 'bad' users. A system to effectively manage reputation is a key issue in the trust of an on-line market place. For example, eBay filed its own reputation based mechanisms for patenting in 2000 [3] and at the beginning of 2015 has almost 200 patents listed on Google's patent with the keyword 'user reputation".

A. Effectiveness of Reputation mechanism

If the reputation mechanism works, known scammers should have the lowest reputation among all users.

Hypothesis 1 Banned users have on average lower reputation than normal users.

If Hypothesis 1 is true, it is evidence that the regulatory mechanism for reputation is effectively enforced, and provides to forum users an instrument to evaluate traders' historical trustworthiness. If the data does not support this, "reputation" in the forum is not a good ex-ante indicator of a users' trustworthiness.

Hypothesis 2 Users with a higher status should on average have a higher reputation than lower status users.

If Hypotheses 1 and 2 do not hold, it may as well be because moderators left a part of the market to its own and concentrated all regulatory efforts on the higher market tiers. For example, in the Carders.de market, there are three Tiers of traders and the first Tier may just represent noise in the data. To check this possibility, we can restrict Hyp1 to holds only for users that are higher in the hierarchy.

Hypothesis 3 Banned users who happened to have a higher status have a lower reputation than other users with the same status.

If even Hypothesis 3 does not hold, we conclude that the reputation mechanisms even after controlling for market alleged statuses provide no meaningful way for the forum users to distinguish between "bad traders" and "good traders".

B. Enforcement of Rules

Reputation may fail to provide effective information, but the hard-wired categories of the forum users (the ones under the direct control of the administrators) may provide a better indicator of quality. Normally, access to the higher market tiers should be subject to some rules.

The market is reliable if such rules are consistently enforced. To see whether this regulation is enforced we can test the following hypothesis:

Hypothesis 4 The ex-ante rules for assigning a user to a Category is enforced.

Once transactions fail, Carders.de and HackMarket.ru users cannot turn to legitimate law enforcement agencies for a redress. Therefore, the forum must have some alternative rules to manage trades gone sour.

Hypothesis 5 There are ex-post rules for enforcing trades contemplating compensation or banning violators.

C. Market Existence

An obvious, but important question to ask is whether the market actually exists. In other words, whether actual transactions take place (took place for Carders.de). Indeed, the role of the forum boards is to provide a platform for sellers and buyers to advertise their merchandise. The actual finalization of the trade usually happens through the exchange of private messages between the trading parties [1], [7].

Hypothesis 6 Users finalize their contracts in the private messages market.

If Hypothesis 6 holds, than the exchange of private messages would be a good proxy for us to measure the successfulness of 'normal' users and 'rippers' in closing trades. To check whether 'normal users' are significantly more successful than 'rippers' we test the following hypothesis.

Hypothesis 7 Normal users receive more trade offers than known rippers do.

For Carders.de, where we have access to the whole forum, a suitable proxy is counting the number of times a forum user initiates a trade with another forum user i.e. the number of unsolicited incoming private messages a user receives. We would expect the results for Hyp. 7 to be coherent with the results obtained so far for the forum. In other words, if the reputation mechanism works, the tier system is properly enforced, and the exchange of private messages is used to conclude the trading process, then we would expect normal users to conclude more trades than rippers do.

COMPARISON OF RESULTS FOR CARDERS.DE AND HACKMARKET.RU.

"Regulation" is the main advantage that a forum-based community has over an IRC-based community: it provides the forum users with a set of rules and mechanisms to assess the

information they can collect on a particular trade. The analyzed markets attempted to enforce this by providing a regulatory mechanism for user reputation and access to “elite” market tiers.

Table I reports the summary of Hypothesis testing for the two markets. The organizational and structural difference of HackMarket.ru with respect to Carders.de is evident.

In Carders.de, each of the regulation mechanisms has been faultily implemented and the potential means for a user to assess ex-ante a trade are pointless or even misleading. The systematic failure of the regulatory mechanisms clearly led to a market where users had no incentives in conducting fair transactions and had no means to distinguish “good traders” from “bad traders”. We showed that there is in fact no difference initiated with a normal user. This effect alone may have brought to the failure of the market, which we show being effectively of the same nature of Florˆencio et al.’s IRC market.

In HackMarket.ru the reputation and punishment mechanisms generate meaningful information for the user:

1) Evidence supports the hypothesis that reputation points are meaningfully assigned to users and this arguably results in a useful tool for the user to assess potential trading partners.

2) The punishment mechanism is a well-regulated one and direct evidence suggests that ‘trials’ are conducted in a fair manner. This boosts market activity and incentivizes ‘honest’ behavior.

3) Users that have been found guilty are, if not banned, publicly exposed and assigned to the ‘scammers’ group. This allows other users to clearly assess a scammer’s trading history and make an informed decision with whom to trade.

Hypothesis	Hyp #	Carders.de	HackMarket.ru
Reputation mechanisms work	Hyp 1	Rejected	Accepted
	Hyp 2	Rejected	N.A.
	Hyp 3	Rejected	N.A.
Regulations are enforced	Hyp 4	Rejected	N.A.
	Hyp 5	N.A.	Accepted
The market works	Hyp 6	Accepted	Accepted
	Hyp 7	Rejected	Accepted

TABLE I: COMPARISON OF RESULTS FOR CARDERS.DE AND HACKMARKET.RU.

Hypotheses aimed at assessing the reliability of the reputation mechanism, the enforcement of regulation, and market fairness are all rejected for Carders.de. In contrast, HackMarket.ru appears to be a well-functioning market.

AVAILABILITY OF EXPLOIT CODE:

The argument against free distribution of exploit code is that "code kiddies" and criminals end up using it [5]. Depriving them of their tools or playthings would obviously make the Internet safer. The argument in favor of exploit code is that it can ensure that a program is secure. This would have arguably been useful in the case of the vulnerability that the Sapphire worm (also known as the SQL Slammer) exploited. Many system administrators who thought they had patched all their systems discovered vulnerable Microsoft SQL servers hidden in many other applications, including some on desktop computers. Yet it isn't clear that system administrators need working exploit code. They could instead use a program that scans for vulnerabilities, such as Nessus.

On the other hand, the Cybercrime Treaty might permit nations to outlaw Nessus and similar scanning programs According to the company; Nessus "will not make its security tests regarding the version number of the remote services, but will really attempt to exploit the vulnerability. An intrusion-detection system would likely flag a Nessus scan from a host outside the network as a probable break-in attempt.

EVILSEED: AN APPROACH

EVILSEED, an approach to search the web more efficiently for pages that are likely malicious. EVILSEED starts from an initial seed of known, malicious web pages. Using this seed, our system automatically generates search engines queries to identify other malicious pages that are similar or related to the ones in the initial seed.

By doing so, EVILSEED leverages the crawling infrastructure of search engines to retrieve URLs that are much more likely to be malicious than a random page on the web. The core of our system is a set of gadgets. These gadgets consume a feed of Web pages that have been previously identified as malicious (as well as other data feeds, such as domain registration feeds). The key idea of our approach is that we can leverage the infrastructure of search engines and the data that they have collected. Of course, the main challenge is to formulate the search queries such that the results indeed have a high probability of pointing to malicious pages.

We assume that we have at our disposal an oracle (and optionally, a pre- filter) that can be used to analyze a web page and determine whether it is malicious or not. Of course, the type of oracle depends on the precise notion of maliciousness that is used. The (evil) seed is a set of pages that have been previously found to be malicious. These pages form the input to gadgets. Of course, whenever gadgets discover new pages that the oracle confirms to be malicious, they can be added to the set of seed pages.

One can distinguish two main types of pages in the seed. First, there are pages that were directly set up by cybercriminals. Typically, these are pages that directly contain scripting (mostly JavaScript) code that launches exploits, or links to malware binaries, such as fake AV programs. A previous paper refers to such pages as malware distribution pages [2]. The second type of pages is not malicious per se. Instead, they are legitimate pages that have been compromised.

Gadget	Expansion	Input
Links	Link topology	Seed URLs, Search Engines
Content dorks	Content similarity	Seed pages source, Search Engines
SEO	Link topology, Seed URLs, Content similarity	Search Engines
Domain registrations	Bulk registrations	Seed URLs, Domain registrations

Table II: GADGETS USED BY EVILSEED.

A. Links Gadget

This gadget is designed to locate “malware hubs” on the web. Malware hubs are pages that contain links to several malicious URLs.1 In our experience, hubs can be grouped in two categories: vulnerable sites that have been infected multiple times (this is typical, for example, of unmaintained web applications), and pages that catalog (and link to) web malware (this is the case of certain malware discussion forums, such as malwareurl.com). This gadget leverages the observation that links contained on malware hubs are likely to be malicious and, thus, represent valuable candidate URLs.

Seed: The seed analyzed by the links gadget consists of all the URLs of known malicious pages.

Expansion: The gadget searches for malware hubs that link to pages in the input seed. More precisely, the gadget issues queries using the link operator, e.g., link :< MALICIOUS URL >. We used multiple search engines to distribute the load of our queries over multiple sites, and to increase the diversity of returned result sets.

B. Content Dorks Gadget

An effective technique to find vulnerable web sites is to query a popular search engine with a Google dork. This term delineates a set of carefully chosen keywords and operators tailored to retrieve links to vulnerable web pages. For example, the query “index of /etc/” will cause the search engine to locate web sites that share their configuration and list of users through their Apache web server. Likewise, a query for “powered by PhpBB 2.0.15” will return web sites that are using an older version of popular bulletin board software with known vulnerabilities.

Seed: As discussed in Section II-B, our initial dataset of malicious pages can be divided into malware distribution pages and landing pages. As a starting point for this gadget, we are interested in landing pages only, which are originally benign but vulnerable pages that have been compromised by an attacker, as opposed to pages directly created by an attacker (e.g., pages generated with an exploit kit).

Expansion: The queries generated by this gadget consist of n-grams of words that are extracted from the indexable content of landing pages in our seed. To generate these n-grams, we use two different techniques: term extraction and n-gram selection.

The term extraction process derives, from the content of a page, those terms that best summarize the topics of this page. This analysis typically leverages techniques from the information retrieval and natural language processing fields. The n-gram selection process extracts all sequences (of length n) of words from a landing page. Then, it ranks all ngrams according to their likelihood of occurring in a malicious page compared to their likelihood of appearing in a benign page. The intuition is that n-grams that appear much more frequently in malicious pages than in benign ones are a good indication for the maliciousness of the page.

C. Search Engine Optimization Gadget

Cybercriminals are able to exploit and take control of large numbers of vulnerable web sites. However, most of these web sites are likely to be part of the “long tail” of the web, and are visited by a very small numbers of users. Therefore, drive-by attacks injected into these websites would only reach a small pool of potential victims. To reach more users, cybercriminals use a variety of techniques to drive traffic to the malicious pages under their

control. Unsurprisingly, these include the use of blackhat Search Engine Optimization (SEO) techniques to increase the ranking of malicious pages in search engine results for popular search terms.

Furthermore, since blackhat SEO campaigns are known to target popular searches obtained from Google and Twitter trends, we extend the seed for this gadget by fetching Google and Twitter trends, querying Google for the resulting topics, and checking the returned URLs with our cloaking detection heuristic.

Expansion: Once we have identified at least one cloaked, malicious URL, we can attempt to locate other URLs in the same blackhat SEO campaign. For this, we use a number of techniques to identify additional candidate URLs. First of all, for each domain hosting a cloaked, malicious web site, we perform a Google query using the site: modifier to locate other pages on that domain. We fetch the query results and add them to the set of candidate URLs. This allows us to find other malicious pages on the same site that may have been optimized for different search terms.

D. Domain Registrations Gadget

Blacklists are one of the most widespread techniques to protect users against web malware. In a domain-based blacklist, a domain is added to the list as soon as it is discovered to host malicious content. As a countermeasure, cybercriminals are known to serve their content from short-lived domains, frequently switching domains to maximize the time they remain unlisted. To run an efficient business in this arms race, cybercriminals are likely to automate the domain generation and registration process.

Seed: The seed used by the Domain Registrations Gadget consists of all the domains that are known to host malicious pages, and domain registration records which are freely available online.

Expansion: This gadget extracts the domain of a malicious seed URL, and flags the domains that have been registered before and after as suspicious.

These domains are then used to create URLs that are scheduled for analysis. The URL creation consists of taking the closest known malicious registration (for which we know the corresponding malicious URL), and replacing the domain with the suspicious domain that we have just flagged.

E. DNS Queries Gadget

The DNS queries gadget analyzes recursive DNS (RDNS) traces. The goal is to identify the domain names of compromised landing pages that are likely to lead to malicious pages. The gadget works in two steps: First, it leverages temporal relationships in a DNS trace to identify domains that are likely connected to malicious domains. More precisely, the gadget checks for queries for domain DL "shortly before" a query for domain DP, which is known to host malicious pages. Then, the gadget identifies pages (URLs) on DL that may redirect their visitors to an evil page P hosted on DP.

Seed: The seed used by the DNS queries gadget consists of all the domains that are known to host malicious pages.

Expansion: This gadget's expansion relies on the fact that, often, a large number of infected pages contain links to a single, malicious page, and that DNS traces (partially) expose these connections. In practice, we passively monitor the recursive DNS traffic generated by a large user base. This traffic can be collected, for example, by deploying a sensor in front of the RDNS server of a network. We assume that a network user, during her regular Internet activity, will browse to a compromised landing page L (hosted on DL) that redirects to P, one of the URLs that are part of our evil seed.

HOST-IP CLUSTERS (HIC):

Since these topologically dedicated hosts and their HICs play a central role in linking different malicious paths together, it becomes important to detect them for breaking the malicious infrastructures. In our research, we come up with a new topology-based technique designed to catch these hosts without relying on the semantics of the attacks they are involved in. Intuitively, these dedicated hosts are rather easy to reach from the dark side of the Web while extremely hard to reach from the bright side. Our approach works surprisingly well: in our evaluation based upon 7-month data crawled from Alexa top websites [1], our approach detects about 5,000 new topologically dedicated malicious hosts and over 20,000 malicious host paths that are not captured by existing solutions, at a false detection rate as low as 2%.

Our major detection results and interesting findings include:

1) Our algorithm achieves a high detection rate. Even with a small set of seed malicious HICs (5% of the labeled ones), we can discover a large number of other malicious HICs, with an expansion rate of 12 times.

2) Our detection algorithm is general across the use of different malicious seeds, including drive-by downloads and Twitter spam in our experiments. It can also detect malicious hosts set up through different attack channels, such as drive by downloads and scam in our data.

3) For the set of dedicated malicious hosts that serve as TDSes, they are much more long-lived than doorways or exploit sites (65 days vs. 2.5 hours). They receive malicious traffic from new attack campaigns over time. Disrupting their operations has more long-lasting effects than taking down do or ways or exploit sites.

4) Our study shows that even after TDSes are taken down, they continue to receive a large amount of traffic, 10 times more than legitimate parked domains. Such traffic is leveraged by domain owners through parking services to continue to gain revenues from ad networks.

DISCUSSION:

The breadth of ICT-enabled criminal activity is shocking [6]. ICTs are regularly being used in the commission of a wide variety of traditional crimes, with serious implications for persons (assault, sex offenses), property (fraud, vandalism) and society (drug offenses, pornography, weapons violations). This calls for a broadening of our understanding about the nature of cybercrime, its sources of risk, and approaches to dealing with it in future. ICTs are also enabling new, unexpected criminal opportunities such as trespassing, identity theft, and denial of service attacks. Existing criminal coding frameworks do not sufficiently capture the prevalence or risks from such emergent technology related crime. Notably, NIBRS lists trespassing as a Group B offense (i.e., an offense that is relatively less serious, frequent, or prevalent nationwide), yet trespassing occurred in over 40% of the cybercrimes cases.

CONCLUSION:

As hackers apply new tactics and seek out increasingly lucrative targets, cyber-attacks and Web scams constitute a significant growing threat to financial firms. From run-of-the- ill email fraud to sophisticated malware infiltration, companies face an ever-widening array of security concerns, the extent of which is often hidden from consumers and shareholders. Although a major trader in grains and feed, Scoular is hardly an industry titan.

The contribution of this paper is twofold. On one side, it replicates and confirms the findings of by showing that a badly regulated cybercrime forum community is virtually no different from an unregulated IRC community.

The second contribution of this article provides an example of regulation in a successful underground community, the (indirect) effects of which are daily reported in security news and industry reports. While the evidence presented in this paper is limited by the scope of the article, it does show that rigorously and well maintained underground markets are possible and do exist. The underground economy should be seen, rather than a confused and unorganized group of criminals scamming other criminals, as a well-organized and administered source of risk that makes for an interesting venue for future research.

ACKNOWLEDGEMENT:

We thank anonymous reviewers for their insightful comments. This work is supported in part by NSF CNS-1223477 and CNS-1117106. Our Department of Computer Science and Engineering also acknowledges the fund from the college B. N. C. O. E. Pusad, India.

REFERENCES

1. J. Franklin, A. Perrig, V. Paxson, and S. Savage, “An inquiry into the nature and causes of the wealth of internet miscreants,” in Proc. Of CCS’07, 2007, pp. 375–388.

2. R. Anderson, C. Barton, R. B¨ohme, R. Clayton, M. van Eeten, M. Levi, T. Moore, and S. Savage, “Measuring the cost of cybercrime.” in Proc. of WEIS’12, 2012.

3. R. Ratterman, R. Maltzman, and J. Knepfle, “Determining a community rating for a user using feedback ratings of related users in an electronic environment,” 2000, US Patent 8,290,809.

4. Federal Bureau of Investigation "National Incident-Based Reporting System," Washington, DC, (Volume 1: Data Collection Guidelines), 2000, pp. 128.

5. L. Lochner "Education and Crime," In *International Encyclopedia of Education*, B. McGraw, P. Peterson and E. Baker (Ed.), Elsevier, Amsterdam, 2008.

6. Hong Kong Police Force. "Technology Crime Division." 2009, accessed Feb 23, 2009 at: www.police.gov.hk/hkphome/english/tcd.

7. C. Herley and D. Florˆencio, “Nobody sells gold for the price of silver: Dishonesty, uncertainty and the underground economy,” in Proc. Of WEIS’10. Springer, 2010, pp. 33–53.

8. K. M. Eisenhardt, “Agency theory: An assessment and review,” Acad. of Manag. Rev., vol. 14, no. 1, pp. 57–74, 1989.