Ask WINE: Are We Safer Today?
Evaluating Operating System Security through Big Data Analysis
Position Paper

Tudor Dumitraș and Petros Efstathopoulos
Symantec Research Labs
{tudor_dumitras, petros_efstathopoulos}@symantec.com

ABSTRACT

The Internet can be a dangerous place: 800,000 new malware variants are detected each day, and this number is growing at an exponential rate—driven by the quest for economic gains. However, over the past ten years operating-system vendors have introduced a number of security technologies that aim to make exploits harder and to reduce the attack surface of the platform. Faced with these two conflicting trends, it is difficult for end-users to determine what techniques make them safer from Internet attacks. In this position paper, we argue that to answer this question conclusively we must analyze field data collected on real hosts that are targeted by attacks—e.g., the approximately 50 million records of anti-virus telemetry available through Symantec's WINE platform. Such studies can characterize the factors that drive the production of malware, can help us understand the impact of security technologies in the real world and can suggest new security metrics, derived from field observations rather than small lab experiments, indicating how susceptible to attacks a computing platform may be.

1 INTRODUCTION

In 2002, researchers demonstrated that a worm can infect all the hosts on the Internet within 30 seconds [6]. In 2004, former hacker Kevin Mitnick estimated that a Windows XP Service Pack 1 system, using the default security settings, can be compromised within 4 minutes after connecting it to the Internet [1]. However, such overwhelming attacks have not been common during the past decade, in spite of an ever growing number of malware strains [7]. Because end users cannot easily determine how safe they are from Internet attacks, and what preventative measures are most effective, in 2012, it is time to ask: *Are we safer today?*

The evolution of Internet attacks has often been described as a type of *arms race* between authors of malicious software and security researchers. For example, today one in 370 emails is a phishing attack, one in 295 emails contains malware, and 2,102 malicious Web-sites and over 800,000 new malware variants are detected each day. Moreover, it is widely assumed that programming errors are unavoidable and that the volume of security vulnerabilities exploited grows with the size of the operating system's code base [4]. Meanwhile, operating systems have evolved as well. Today, they have reduced attack surfaces—the number and importance of attack vectors that viruses have at their disposal (e.g., open sockets that may receive a buffer overflow exploit) [3]—and security updates patching vulnerabilities are released frequently. For example, during the last 10 years several security technologies have been added to the Windows operating system, such as software data execution prevention (safe structured exception handling), hardware data execution prevention (the NX bit), or address space layout randomization (ASLR). Many of these technologies are able to protect not only the operating system, but also the running programs.

To understand which side is winning this arms race, it is important to understand and *measure the effectiveness of cyber attacks*. One resource available for conducting such a study is the Worldwide Intelligence Network Environment (WINE), a platform for data intensive experiments in cyber security [2]. WINE was developed at Symantec Research Labs for sharing comprehensive field data with the research community. WINE samples and aggregates multiple terabyte-size data sets, which Symantec uses in its day-to-day operations, with the aim of supporting open-ended experiments at scale. WINE also enables the reproduction of prior experimental results, by archiving the reference data sets that researchers use and by recording information on the data collection process and on the experimental procedures employed. For example, one of the WINE data sets includes telemetry received from Symantec's anti-virus products, collected from 0.5–1 million Windows hosts per month for nearly two years. These hosts do not represent honeypots or machines in an artificial lab environment; they are real computers, in active use around the world, that are targeted by cyber attacks.

We have two goals: (i) to verify or disprove popular beliefs about viruses and malware (evolution, spreading, infection patterns etc.); and (ii) to investigate correlations between security features of an operating system and its observed susceptibility to attacks. This study is part of a

broader research agenda aiming to understand cyber security through the analysis of comprehensive field data, made available through WINE. The remainder of this paper is organized as follows: in Section 2 we describe the data available and how we can analyze it. In Sections 3 and 4 we describe the research opportunities for our two goals and provide preliminary results, and in Section 5 we summarize our conclusions and discuss our plans for future work.

2 ANALYZING A/V TELEMETRY

The anti-virus telemetry data set in WINE records occurrences of known threats (e.g., worms, viruses, trojans), detected by Symantec's anti-virus products on end-hosts around the world. Each detection indicates that the anti-virus has stopped an attack that may have resulted in an infection. Therefore, this data set does not provide any information on successful infection rates or on the vulnerabilities present on the targeted hosts. Rather, as these detections correspond to malicious files that are downloaded or stored on these hosts, having bypassed the operating system's defenses, the data allows us to measure how susceptible end-hosts are to attacks [1]. This is similar to the concept of *attack surface* [3], which can be measured—without making assumptions about unidentified vulnerabilities in the system or about the attacker's capabilities— by considering a system's attack opportunities, such as open ports or services running by default. Unlike in measurements of the attack surface, we focus on actual *attacks observed in the real world*, rather than on opportunity estimations.

The data collection focused on the versions of the Windows operating system, which is currently the primary target for malware attacks [7]. As with any field-gathered corpus of data, some records are incomplete, inaccurate or difficult to interpret. For example, we can extract from each record the Windows version, the Service Pack information and the build number. Some combinations correspond to pre-RTM builds (e.g., the 71xx and 72xx build numbers) and some do not correspond to actual Windows releases (e.g., "Dear build 7600"). A preliminary analysis suggests that platforms detected on at least 1,000 hosts cover versions from Windows 2000 SP4 to Windows 7 SP1, spanning 9 years of Windows releases. A more challenging problem is to determine the number of distinct viruses that target these operating systems. Anti-virus signatures are often generic, covering a range of malware samples, they employ heuristics and they sometimes look for exploits that may be re-used by multiple viruses. In general, because it is easy to produce new malware strains based on the code of older viruses, the precise difference between a virus and a virus family is subject to interpretation. One possible approach would be to count only the signatures that do not detect files which are also detected by other signatures (i.e., the detections that do not overlap).

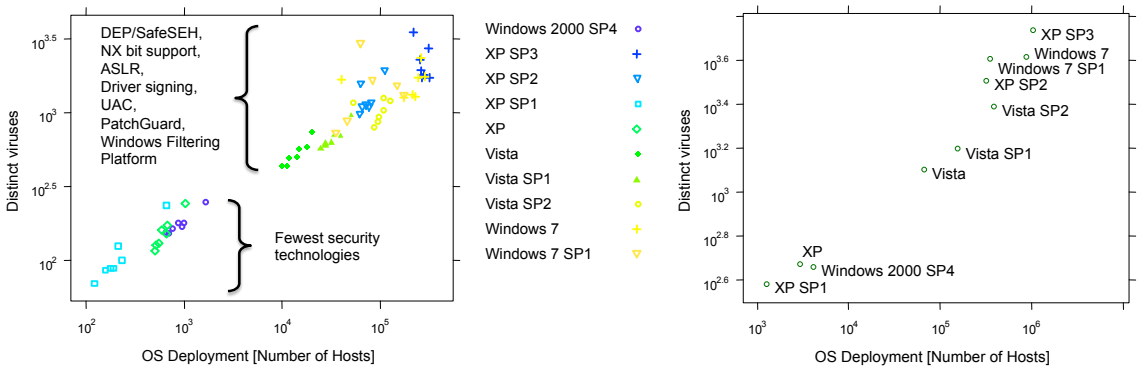
Threats to validity. The biggest threat to the validity of this study is selection bias. As WINE does not include telemetry from hosts without Symantec's anti-virus products, our results may not be representative of the general population of platforms in the world. In particular, users who install anti-virus software might be more careful with the security of their computers and, therefore, might be less exposed to attacks. Although we cannot rule out the possibility of selection bias, the large size of the population in our study and the diversity of platforms suggest that our results have a broad applicability.

3 MEASURING THE CYBER THREAT LANDSCAPE

In our first analysis, we focus on quantifying the evolution of the threat landscape. In particular, one of our goals is to assess the growth rate of viruses and malware that are being observed on the field, and to assess common assumptions about the evolution of cyber threats, based on field data. As malware authors have historically focused on the Windows platform, which is the most prevalent around the world, we expect to see a similar trend among Windows versions—prioritizing malware development for the platforms with the widest deployment. This is illustrated in Figure 1a. During our data collection period, Windows XP SP3 has been the most widely deployed platform with Windows 7 being a close second. As observed in Figure 1a (as well as Figure 1b, included for clarity) both of these OS versions were affected by the largest sets of viruses. In fact, the number of distinct viruses that affect a platform increases linearly with the number of active hosts using this platform around the world. We note that this trend holds in spite of changes in the deployment of different platforms by up to one order of magnitude during the observation period.

Linking the threat landscape growth with platform deployment size reveals an important correlation, that can guide our expectations regarding the threats. However, we believe that by further analyzing anti-virus telemetry, as well as other data-sets at our disposal, we can identify more platform properties that can be used as a (cumulative) field metric for threat landscape evolution.

[1] Notice that estimating end-host susceptibility to attacks based on the amount of reported detections is a heuristic, and hence not meant to be absolutely accurate. In particular, one might claim that operating systems defences might block the attack if the anti-virus software had not done so ahead of time. However, from a security standpoint, this is a very risky assumption make. Furthermore, field data suggest that in a large number of cases this is not true. In any case, the mere presence of a threat on a machince suggests some degree of (perceived) susceptibility to attacks, making us feel confident that our heuristic is based on a fair and realistic assumption.



(a) Correlation between the deployment size of Windows platforms and the number of distinct viruses detected on them, during a 6-month period starting in March 2011.

(b) Correlation between the deployment size and the number of distinct viruses, aggregated for the entire period.

Figure 1: Impact of a platform's size of deployment on the production of malware targeting the platform. The data is plotted on a logarithmic scale to emphasize the trend for platforms with small deployments.

4 A/V TELEMETRY: IMPACT OF SECURITY TECHNOLOGIES

Aside from being a good indicator of threat landscape growth, anti-virus infection reports are, by definition, one of the most indicative data sources regarding the security state of the hosts, and their susceptibility to attacks. During the second part of our study we intend to investigate methods that can help us assess a host's vulnerability level, based solely on telemetry data, and in particular infection reports. This is part of our greater effort to analyze WINE data, so as to discover significant correlations allowing us to define security metrics that can be applied directly to field data, and provide (i) statistically reliable (yet not necessarily 100% accurate) estimates of host security properties, and (ii) alternatives to other such metrics, that can be applied directly to real-world data with, potentially, small experimental investment and/or run-time cost. In this context, in particular, we aim to extrapolate end-point susceptibility to attacks from anti-virus detection reports. This would allow us to express an informed opinion that can be used a standalone measure of a particular platform's security status, or as complementary input to existing/alternative methods with similar goals—such as Microsoft's Attack Surface Analyzer [5]. For instance, these results can be used either as a focusing method for other tools (e.g., determine problematic platforms for attack surface analysis), or as a "post" evaluation tool—measuring the effectiveness of newly introduced features and/or inspection techniques.

With regard to the evaluation of newly introduced technologies, we are particularly interested in investigating the relationship between particular security technologies introduced in Microsoft Windows, and their effect on the number of observed attacks. Data execution prevention (DEP) via Safe exception handling, address space layout randomization (ASLR), NX-bit support, new user account management techniques, the Windows filtering platform, driver signing, and the Patchguard technology are some key security technologies that have been introduced in different releases and Service Packs of Microsoft Windows. By identifying the point of introduction of each mechanism in our data set timeline, we can study their real-world impact on platform susceptibility to attacks and provide a measure of their effectiveness.

Revisiting Figure 1a, we can observe that data points form two clusters, each containing one or more OS versions, as depicted more clearly in Figure 1b. These clusters loosely correspond to the major security technology improvements introduced in different version of Windows. The first cluster, in the bottom left corner, consists of Windows versions (Windows 2000 SP4, Windows XP and Windows XP SP1) that still have a significant user-base (and hence attract a substantial number of threats) but are being slowly phased-out. These platforms incorporate a variety of improvements, but none of the important security mechanisms mentioned above. The top right cluster consists of the more popular Windows versions that incorporate some or all of the latest security technologies. Windows XP SP2 was a major security release, by incorporating an improved firewall that was enabled by default, NX-bit support and Patchguard, while Windows Vista essentially marked a new era by incorporating all of the above mentioned technologies. Nevertheless, the wide deployment of the OSes in the third cluster accounts for the large number of attacks against them.

Furthermore, recent security incidents have been contributed to unsuccessful updates and flawed software patches—thus generating reluctance among IT profes-

sionals regarding OS upgrades. The correlation of major and minor OS releases with large amounts of field data (including but not limited to anti-virus reports) aims to provide additional metrics to evaluate susceptibility to attacks, by gaining a macroscopic view of the effectiveness of Service Packs and major operating system releases, with respect to security. We believe that performing this type of analysis on millions of real-world security telemetry reports, will help us determine whether we can fairly consider end-points to be safer today, and quantify our claims.

5 CONCLUSIONS AND FUTURE WORK

In this paper, we propose a research agenda for evaluating operating system security through the analysis of comprehensive field data. As a concrete example, we discuss the new knowledge that can be derived from a data set of anti-virus telemetry, collected from 0.5–1 million Windows hosts per month for nearly two years. This data is available to the research community through Symantec's WINE platform for data-intensive security experimentation, in order to identify important correlations between measurable security properties and platform features. This research direction will allow us to define new security metrics that can be derived directly from extensive observations of attacks in the real world, rather than from estimations of attack opportunities.

To illustrate the potential of this work-in-progress, we have also present some preliminary results. For instance, we were able to produce quantifiable evidence of the threat landscape growth, by determining that the number of distinct viruses observed in the field for each OS version is correlated with factors that describe the target of opportunity for cyber criminals, such as the version's deployment size. Furthermore, we have devised a concise plan to analyze anti-virus reports with respect to OS major and minor release history, particularly in connection with the introduction of major security technologies.

As part of our future work, we first plan to perform the anti-virus data analysis described in this position paper. Furthermore, we want to correlate our findings with other data-sets in WINE, such as field data about system and application crashes. Finally, we would like to investigate ways to use our results in order to complement existing methods with similar goals—such as attack surface measurement schemes.

REFERENCES

[1] L. Cheng. Automated "bots" overtake PCs without firewalls within 4 minutes, Nov 2004. Retrieved on 26 Feb 2012.

[2] T. Dumitraş and D. Shou. Toward a standard benchmark for computer security research: The Worldwide Intelligence Network Environment (WINE). In *EuroSys BADGERS Workshop*, Salzburg, Austria, Apr 2011.

[3] M. Howard, J. Pincus, and J. M. Wing. Measuring relative attack surfaces. In *Workshop on Advanced Developments in Software and Systems Security*, Taipei, Taiwan, Dec 2003.

[4] G. McGraw. *Software Security: Building Security In*. Addison-Wesley, 2006.

[5] Microsoft Corporation. Microsoft Attack Surface Analyzer - Beta. `http://www.microsoft.com/download/en/details.aspx?id=19537`.

[6] S. Staniford, V. Paxson, and N. Weaver. How to 0wn the internet in your spare time. In *Proceedings of the 11th USENIX Security Symposium*, pages 149–167, Berkeley, CA, USA, 2002. USENIX Association.

[7] Symantec Corporation. Symantec Intelligence Report: January 2012. `http://www.symanteccloud.com/en/us/mlireport/SYMCINT_2012_01_January_FINAL-en.pdf`.