WHITE PAPER

Unlicensed Software and Cybersecurity Threats

Sponsored by: BSA | The Software Alliance

John F. Gantz
Thomas Vavra
Victor Lim
Pavel Soper
Professor Lars Smith (University of Louisville)
Stephen Minton

January 2015

INTRODUCTION

This White Paper analyzes the relationship between unlicensed software use and malware encounters, and it draws three conclusions: first, that there is a strong correlation between the two variables; second, that unlicensed software use is a strong *predictor* of malware encounters; and third, that there is empirical evidence of causation.

Analysts have long been aware that there is a connection between unlicensed software and cybersecurity threats. For example, when the "Conficker" worm spread through computers around the world in 2008 and 2009, security analysts warned that downloading unlicensed software was among the likeliest ways to get infected.[1] A few years later, the takedown of the Citadel botnet – which created 5 million zombie computers across 90 countries – revealed that the criminals behind it had infected PCs in part by selling unlicensed versions of Microsoft Windows pre-infected with Citadel malware.[2] So it came as no surprise when the FBI issued a consumer alert in 2013 warning that unlicensed software may contain malware.[3]

But there has not yet been a thorough statistical analysis of the connection between unlicensed software and security threats from malware. Accordingly, BSA | The Software Alliance asked IDC to examine the evidence. The findings of this analysis strongly suggest that public policies and firm-level best practices that ensure software is properly licensed will contribute to more secure computing environments.

DETERMINING CORRELATION

To address the connection between unlicensed software and security threats, IDC analyzed rates of unlicensed software use and cybersecurity threats in 81 countries where authoritative data are available on both.

[1] See the June 20, 2011, Krebs on Security blog post entitled "Software Cracks: A Great Way to Infect Your PC" and related comments at http://krebsonsecurity.com/2011/06/software-cracks-a-great-way-to-infect-your-pc/.

[2] A short write-up about the Citadel takedown can be found on the BBC News Web site in a June 6, 2013, article entitled "FBI and Microsoft take down $500m-theft botnet Citadel." See http://www.bbc.com/news/technology-22795074.

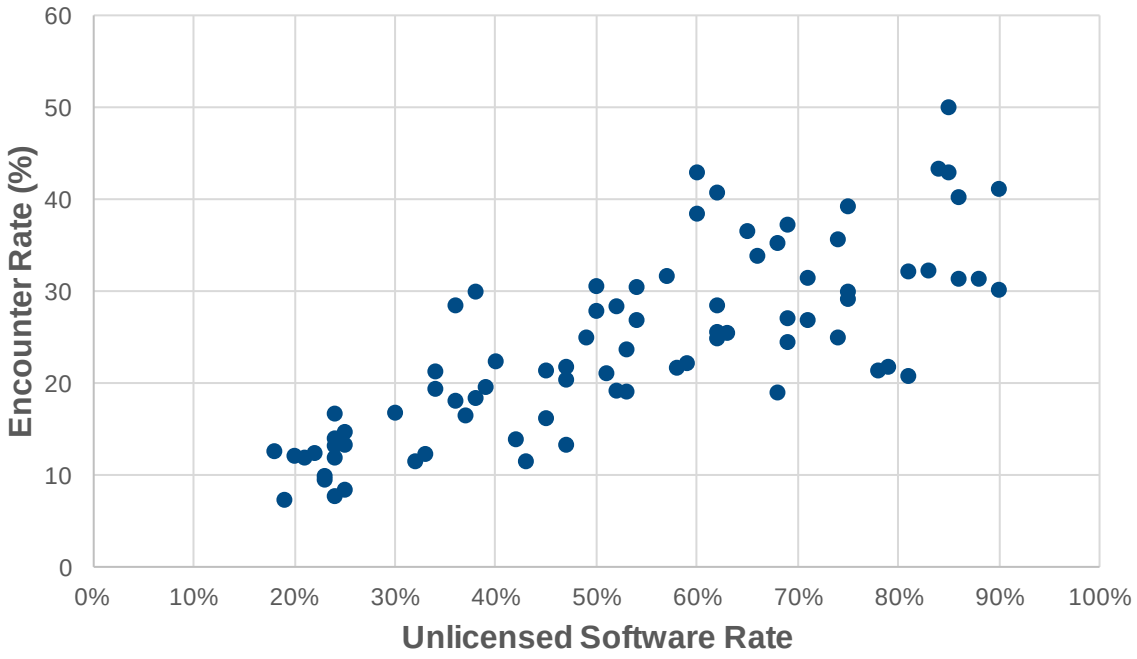
[3] Issued August 2013 and available at http://www.fbi.gov/news/stories/2013/august/pirated-software-may-contain-malware.

Unlicensed software rates come from the *Global Software Survey,* a biennial study that IDC conducts for BSA,[4] and cybersecurity threat information comes from Microsoft's Security Intelligence Report,[5] which looks at activity on 600 million users' computers per month. The metric chosen for the purposes of this White Paper was the *encounter rate,* which is the percentage of computers running Microsoft real-time security software that report detecting malware in a quarter. For perspective, about 20% of PCs worldwide reported malware encounters each quarter in 2013.[6]

Figure 1 shows the data points for both the rate of unlicensed software use and the prevalence of malware encounters in each of the 81 countries for which both encounter rates and unlicensed software rates were available in 2013.

FIGURE 1

Unlicensed Software Rates and Malware Encounter Rates Are Strongly Correlated



Each dot represents an individual country's rate of unlicensed software use and prevalence of malware encounters. (See the Appendix for complete data.) The pattern represents a statistically strong correlation of 0.79 between the two variables.

Source: IDC, 2015

[4] *BSA Global Software Survey: The Compliance Gap,* June 2014, available at http://www.bsa.org/studies.

[5] See Volumes 15 and 16, for descriptions of the data and methodology, available at http://www.microsoft.com/security/sir/default.aspx.

[6] Is this the best measure of cybersecurity threats? There are others, published by companies like Cisco, IBM, Kaspersky, Microsoft, Symantec, Trend Micro, and Verizon as well as government and computer emergency response teams, but most, if they have country-specific information at all, look at threat sources, not destinations. Using a metric designed for PCs and tracked across many countries is also appealing in comparing to a metric based on PC software usage.

The values clearly trend upward together: the higher the unlicensed PC software rate in a country, the more malware generally encountered on PCs in that country.

For example, in 2013, the unlicensed software rate for the United States was 18% and the malware encounter rate averaged 13% per quarter. For Indonesia, the unlicensed software rate was 84% and the malware encounter rate averaged 44% per quarter. Brazil, with an unlicensed software rate of 50%, had a malware encounter rate of 31% per quarter.

Statistical analysis confirms that the two sets of variables have a strong positive correlation, meaning they move up and down together. The correlation coefficient in this case is 0.79, where 1.0 represents a perfect correlation and 0 represents no correlation. By comparison, the correlation between smoking and lung cancer is 0.72,[7] the correlation between education and incomes is 0.77,[8] and the correlation between anti-corruption policies and economic growth is 0.77.[9]

While this correlation neither proves nor disproves causation, it clearly shows that when unlicensed software rates are lower, malware encounter rates also are lower.

BUILDING A PREDICTIVE MODEL

The next step in the analysis was to develop a model to show how accurately unlicensed software rates could be used to *predict* malware encounters. The authors did this with a statistical technique called regression analysis, which involves using the data sets to derive a formula by which one variable (the rate of unlicensed software use) can predict another (the malware encounter rate).

Figure 2 shows the results of that analysis. If the formula worked perfectly, all of the values would be on the line. If the formula didn't work at all, the values would be scattered randomly. In this case, most of the values are clustered near the line, with a statistically strong predictive value (known as R-squared) of 0.62 – meaning the model worked quite well. It can be interpreted that 62% of the variability between one country's malware encounter rate and another's can be attributed to the variability in the respective unlicensed software rates of those countries.

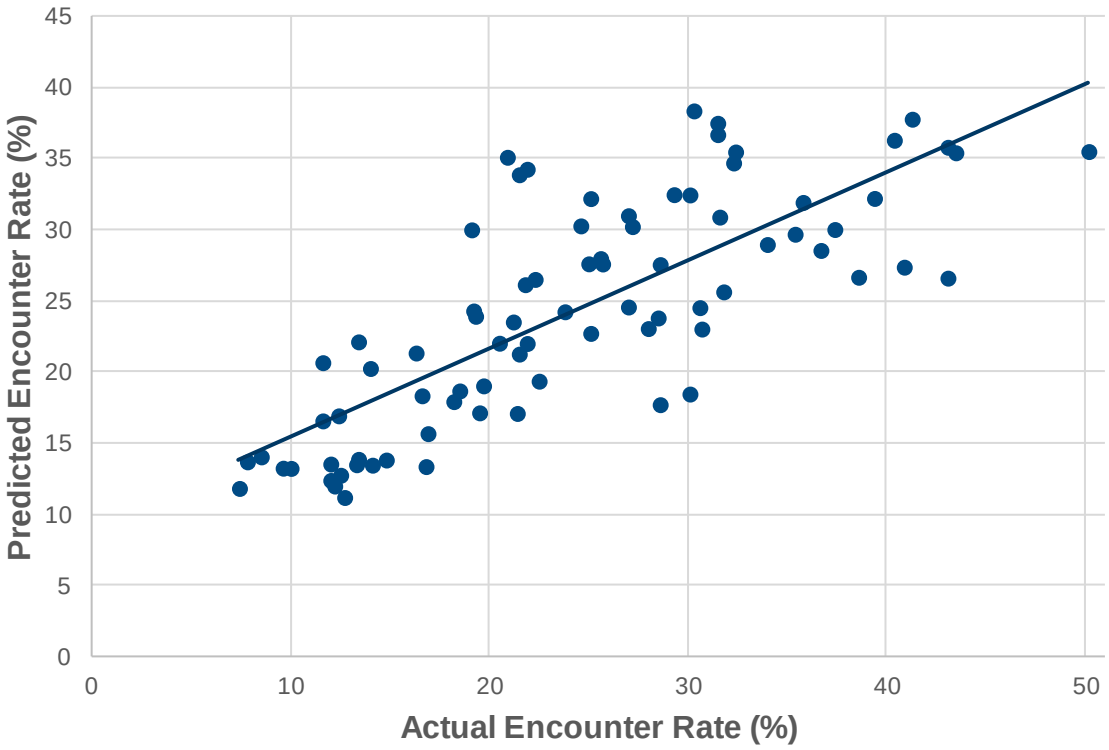
[7] This was a government study in England looking at the number of cigarettes smoked per day compared to lung cancer rates among thousands of men in 25 occupational groups. An abstract can be found at http://www3.nd.edu/~busiforc/handouts/Data%20and%20Stories/correlation/smoking%20and%20cancer/smoking.html, and a detail of the calculations used to get the correlation coefficient can be found at http://www.spcforexcel.com/correlation-analysis.

[8] International Education Statistics, by Friedrich Huebler, August 2008, available at http://huebler.blogspot.com/2005/09/national-wealth-and-school-enrollment.html. Pearson correlation by IDC.

[9] See the OECD issues paper on corruption and economic growth at http://www.oecd.org/g20/topics/anti-corruption/issues-paper-on-corruption-and-economic-growth.htm.

FIGURE 2

Unlicensed Software Use Is a Strong Predictor of Malware Encounters



Each dot represents an individual country's rate of unlicensed software use and predicted rate of malware encounters. (See the Appendix for complete data.) The pattern shows a statistically strong predictive value (R-squared) of 0.62 between unlicensed software use and malware encounters.

Source: IDC, 2015

EVIDENCE OF CAUSATION

It may come as little surprise that unlicensed software use and malware encounters are highly correlated or that a regression analysis finds that one strongly predicts the other. On their own, however, these findings do not prove that lowering unlicensed software rates also would lower malware encounter rates. To reach that conclusion, one must view the statistical analysis in the context of the fact that there is strong *empirical* evidence of a causal relationship.

To put this in context, two variables can very easily have a high correlation value but a low predictive value in a regression analysis. It occurs when the correlation is mere coincidence. For example, it has long been noted that there is a high correlation between ice cream sales and murder rates in the

United States, and it seems obvious that one doesn't cause the other (although hot weather may cause both).[10] Here, however, there *is* causal evidence.

For example, a 2014 study conducted by IDC and the National University of Singapore (NUS)[11] revealed significant amounts of malware in unlicensed software across more than 800 tests of PCs purchased with unlicensed software pre-installed, of unlicensed software DVDs, and of unlicensed software and activation keys downloaded from the Internet. The tests spanned a dozen countries across Asia, Europe, and the Americas. Their conclusion: On average, a user of an unlicensed software package has a one-in-three chance of encountering malware.

This infection rate multiplied by the number of unlicensed software packages in the world suggests there are in excess of 500 million infected unlicensed software packages in circulation. (The research also found that more than 40% of consumers did not routinely install automated security updates, which can also enable malware infections of PCs.)

In a corresponding survey of nearly 1,000 PCs from 15 countries, the IDC-NUS study found that 1 in 5 respondents said that unlicensed software infected their PCs with a virus; 2 in 5 said it substantially slowed their computers and had to be uninstalled (a possible sign of hidden malware); and 1 in 10 said it destroyed files.

Given such evidence, it is not surprising that BSA's 2013 *Global Software Survey* found that computer users around the world cite exposure to security threats from malware as the chief reason *not* to use unlicensed software.

CONCLUSION

This statistical analysis and evidence from the field point to a clear link between unlicensed software and cybersecurity threats. Not all cybersecurity threats come from malware, and not all malware comes from unlicensed software – and *most* malware constitutes a cybersecurity threat.[12]

For enterprises, governments, and consumers, the obvious implication is that one way to lower cybersecurity risks is to reduce the use of unlicensed software. Doing so requires implementing effective software management policies and procedures and investing resources in increasing awareness of the potential dangers associated with using unlicensed software. The dangers lurk in

[10] Justin Peters of *Slate* has summarized the reporting on this particular correlation. See "When Ice Cream Sales Rise, So Do Homicides. Coincidence, or Will Your Next Cone Murder You?" July 9, 2013, at http://www.slate.com/blogs/crime/2013/07/09/warm_weather_homicide_rates_when_ice_cream_sales_rise_homicides_rise_coincidence.html.

[11] *The Link between Pirated Software and Cybersecurity Breaches,* published in March 2014. It is available at http://news.microsoft.com/download/presskits/dcu/docs/idc_031814.pdf. This study followed previous studies by IDC published in 2013 and 2007 on malware in unlicensed software.

[12] In its *2013 Data Breach Investigations Report,* Verizon found that 40% of threat events involved malware and that 71% targeted end-user devices. See http://www.secretservice.gov/Verizon_Data_Breach_2013.pdf.

malware that can be embedded in the software, in the sites and sources from which the malware is obtained, and in the reluctance of users of unlicensed software to install security updates. But the evidence shows that unlicensed software is clearly associated with security threats from malware – the global costs of which run into the hundreds of billions of dollars a year.[13]

[13]See *The Link between Pirated Software and Cybersecurity Breaches*, op. cit.

APPENDIX — COUNTRIES AND DATA USED IN THIS STUDY

Table 1 contains a list of the countries and data included in this study.

TABLE 1

Unlicensed Software Rate and Malware Encounter Rate by Country, 2013 (%)

Country	Unlicensed Software Rate	Malware Encounter Rate
Moldova	90	30
Georgia	90	41
Venezuela	88	32
Belarus	86	32
Iraq	86	40
Algeria	85	43
Pakistan	85	50
Indonesia	84	44
Ukraine	83	32
Nigeria	81	21
Vietnam	81	32
Guatemala	79	22
Kenya	78	22
Albania	75	29
Dominican Republic	75	30
Tunisia	75	39
China	74	25
Kazakhstan	74	36
Lebanon	71	27
Thailand	71	32
Argentina	69	25
Serbia	69	27
Philippines	69	37
Uruguay	68	19
Ecuador	68	35

TABLE 1

Unlicensed Software Rate and Malware Encounter Rate by Country, 2013 (%)

Country	Unlicensed Software Rate	Malware Encounter Rate
Morocco	66	34
Peru	65	37
Bulgaria	63	26
Greece	62	25
Romania	62	26
Russia	62	29
Egypt	62	41
India	60	39
Turkey	60	43
Chile	59	22
Kuwait	58	22
Jordan	57	32
Malaysia	54	27
Mexico	54	31
Latvia	53	19
Lithuania	53	24
Croatia	52	19
Colombia	52	29
Poland	51	21
Saudi Arabia	50	28
Brazil	50	31
Qatar	49	25
Estonia	47	13
Cyprus	47	21
Italy	47	22
Slovenia	45	16
Spain	45	22
Hong Kong	43	12
Puerto Rico	42	14

TABLE 1

Unlicensed Software Rate and Malware Encounter Rate by Country, 2013 (%)

Country	Unlicensed Software Rate	Malware Encounter Rate
Portugal	40	23
Hungary	39	20
Taiwan	38	19
Korea	38	30
Slovakia	37	17
France	36	18
UAE	36	29
Czech Republic	34	20
South Africa	34	21
Ireland	33	12
Singapore	32	12
Israel	30	17
Norway	25	9
Canada	25	13
Netherlands	25	15
Finland	24	8
Switzerland	24	12
Germany	24	13
United Kingdom	24	14
Belgium	24	17
Sweden	23	10
Denmark	23	10
Austria	22	13
Australia	21	12
New Zealand	20	12
Japan	19	7
United States	18	13

Source: IDC, 2015

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

5 Speen Street
Framingham, MA 01701
USA
508.872.8200
Twitter: @IDC
idc-insights-community.com
www.idc.com