Cybersecurity Workload Trends

DAN GEER AND ERIC JARDINE

Dan Geer is the CISO for In-Q-Tel and a security researcher with a quantitative bent. He has a long history with the USENIX Association, including officer positions, program committees, etc. dan@geer.org

Eric Jardine is an Assistant Professor of Political Science at Virginia Tech and a Fellow at the Centre for International Governance Innovation. His research focuses on issues to do with measurement and cybersecurity, the uses and abuses of the Dark Web, and trust and the Internet ecosystem. ejardine@vt.edu.

> I became interested in long-term trends because an invention has to make sense in the world in which it is finished, not the world in which it is started.—*Ray Kurzweil*

A small bit of statistical wisdom: trend analysis can derive real guidance even when the measurement being examined is subject to consistent (relatively constant) error. Hold that thought...

NIST (the US National Institute of Standards & Technology) has for years collated and published vulnerability information, with the Common Vulnerability Scoring System (CVSS) being the best known of NIST's cybersecurity metrics. CVSS scores are numeric and calculated by a defined, constant formula [1]. Putting aside that calculation formula, CVSS is a stable system for which the errors are relatively constant.

From the CVSS data, NIST publishes on a daily basis what it calls a Workload Index, defined this way [2]:

> This [Workload Index] calculates the number of important vulnerabilities that information technology security operations staff are required to address each day. The higher the number, the greater the workload and the greater the general risk represented by the vulnerabilities.
>
> The NVD workload index is calculated using the following equation:
>
> (
> (number of high severity vulnerabilities published within the last 30 days) +
> (number of medium severity vulnerabilities published within the last 30 days/5) +
> (number of low severity vulnerabilities published within the last 30 days/20)
>) / 30
>
> The index equation counts five medium severity vulnerabilities as being equal in weight with 1 high severity vulnerability. It also counts 20 low severity vulnerabilities as being equal in weight with 1 high severity vulnerability.

Taking the Workload Index to be, just as it says, a composite estimate of the workload imposed on information technology security operations staff by the changing inventory of vulnerabilities in the CVSS catalog, we can begin to ask some questions.

The first and most obvious would be simply whether the workload due to known vulnerabilities is improving (going down) or worsening (going up). In finance, a typical measure of how a company is doing is "trailing twelve month" income—the income for the twelve-month period immediately prior to the date of the report. In Figure 1, we show the trailing 12-month value of the Workload Index over the past decade (overlain with a fitted order-2 polynomial, and with the X axis crossing the Y at Y=0).

Does that curve tell us anything? It certainly appears that information technology security operations staff had a few years of declining workload but may now be in a period of rising workload. One almost imagines a suite of arguments paralleling those about global warming to break out here—is workload rising or is this just natural variation?

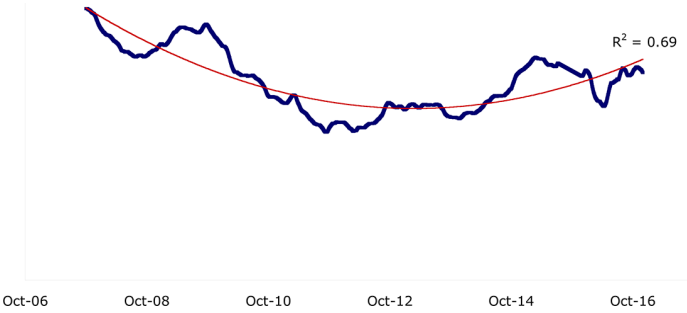

Figure 1: Trailing 12-month Workload Index

In finance, the measure of variation is called "volatility," usually expressed as the trailing 12-month standard deviation. So, in Figure 2 we show exactly that, the trailing 12-month standard deviation of the Workload Index (again overlain with a fitted order-2 polynomial, and with the X axis crossing the Y at Y=0).

We might now ask (ourselves) how strong is the indication that volatility in the Workload Index is rising? Nassim Taleb, whom you may know from having read some of his *Incerto* tetralogy [3], has characterized a system with rising interconnectedness as one where a "black swan" event can (will) occur. In particular, he suggests that our hyper-connected society is "undergoing a switch between [continuous low grade volatility] to ... the process moving by jumps, with less and less variations outside of jumps." The NVD Workload Index cannot itself answer a conjecture that

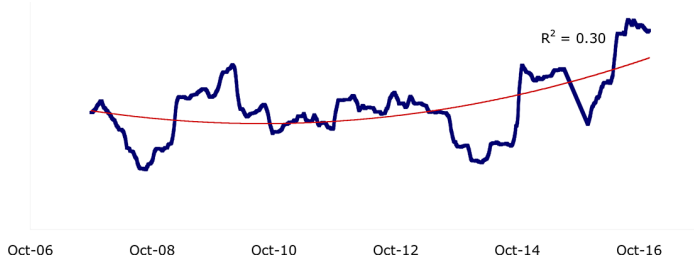

Figure 2: Trailing 12-month standard deviation (volatility)

serious vulnerabilities are becoming rarer except for the few that slip through and are found to be more serious than ever. What do you see in Figure 2?

So what is the meaning of "workload" anyhow? Can we think of it as interest on technical debt? Does it need some sort of normalization to be a worthy basis for decision-making? There is no doubt that the source of risk is dependence, particularly dependence on the stability of system state, so is this workload measure, along with other measures, a way to price our dependence? Or is it something else?

Let's think first about economy-wide effects. The number of schools offering instruction in cybersecurity has skyrocketed in the last decade [4]. All those people entering the field should have the effect of divvying up the workload, shouldn't they? The Index of Cyber Security [5] looked at one form of that question, asking it twice, 40 months apart: "As you look to fill vacancies in your organization, which of the following describes the status of the current job market for information security professionals?"

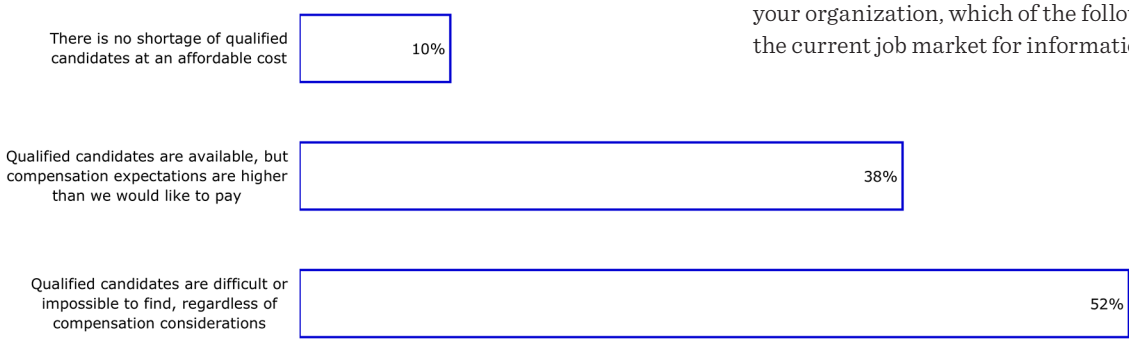

Figure 3A: November 2012

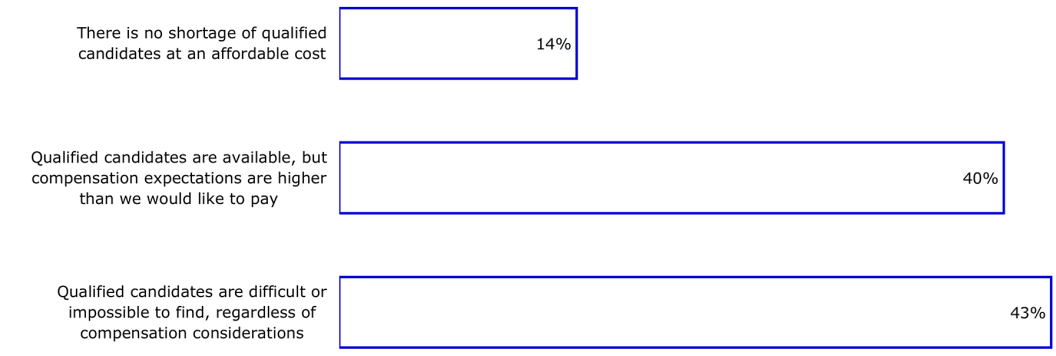

Figure 3B: April 2016

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
mean WI	10.79	9.59	8.69	8.96	5.99	5.98	6.51	6.05	7.88	7.65
100K workers	4.01	4.67	4.75	4.71	5.37	5.53	6.05	6.02	6.29	6.52
WI/100K	2.69	2.05	1.83	1.90	1.12	1.08	1.08	1.01	1.25	1.17

Table 1: Workload Index normalized by number of workers

2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
54,111	47,480	42,170	38,476	37,992	39,593	43,066	47,406	50,962	55,367

Table 2: Bachelor's degrees in computer and information sciences

Figure 3A shows the answer in November 2012, while Figure 3B shows the answer in April of 2016.

From the first sample in 2012 to the second in 2016, the idea that "qualified candidates are difficult or impossible to find" fell by almost 10 percent. The answer that those frontline security managers gave implies an increasing supply of competent individuals with whom to share the workload. Can we normalize to that? And if we do, might that tell us more about the level of cybersecurity risk from technical vulnerabilities in the economy?

Table 1 shows the yearly average Workload Index number from 2006 to 2015, which can, in turn, be normalized by the US Bureau of Labor Statistics dataview for the number of workers in the category "Computer and information systems managers" [6]:

The data in Table 1 is redrawn as a chart in Figure 4, again overlain with a fitted order-2 polynomial. If you imagine plotting the mean Workload Index onto Figure 4 as well, you would have a line that declines into 2011, but then increases a fair amount from there on in. In this case, we see a steady decline and flattening of the curve when the index is normalized to the number of workers. Framed in this light, the "workload" posed by new vulnerabilities has gotten better since 2006 and remained relatively flat ever since. (Note that BLS data for the preferred category "Information security analysts" only began in 2011, so that category cannot yet be used for decadal views.)

Managing a variable amount of risk in a large system is only partially about the particular risks currently in that system; it is about the history (and future) of scaling factors as well. Sometimes, from 2006 to 2011, for example, when the mean score on the Workload Index was declining, one might naturally have inferred that cyberspace was becoming safer. Should we now infer that that welcome decline has stopped?

Over the last decade, the number of new graduates entering the workforce with computer science degrees fell and then rose, as seen in Table 2.

Those annual graduation numbers, as it turns out, are not correlated with the numbers of "Computer and information systems managers" in the workforce (r = .18), so either there is a lot of turnover among those jobs or the graduates are going somewhere else. So we will stick with "Computer and information systems managers" as our description of who is handling the vulnerability workload. But the Workload Index is really about how much work there is to be done. If we think of the work to be done as handing each member of the workforce a to-do list, then we would multiply the workforce count by the Workload Index and call that a measure of the work pending in the economy at large, viz., the size of the to-do list in the economy at large. That gets you Figure 5.

This mathematical manipulation generates an economy-wide to-do list, but labor markets can be sticky, as evidenced by the lack of a correlation between new computer science graduates and computer and information systems managers. This means that the "real" level of risk in the system might not have translated over into enough workers to actually handle the daily updates and patches needed to address the Workload Index. In

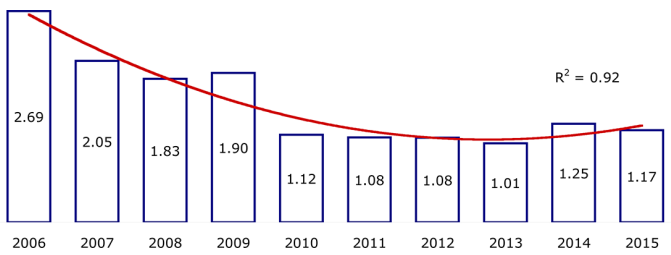

Figure 4: Workload Index normalized to number of workers

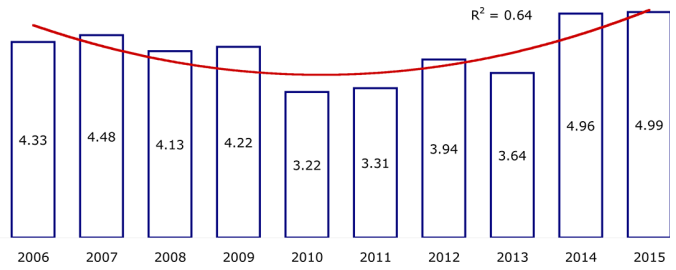

Figure 5: Millions of Items on the economy-wide to-do list

such a situation of a labor market failure, the formula for the economy-wide to-do list would be something like Y = Workload Index * (computer and information systems managers + X), where X is the number of workers who should be working in the system but are not due to market lags. X, in a world of imperfect information and with humans who need to be educated and trained, would be some positive value—at least until we swing past the labor market saturation point into a surplus labor situation.

What would this do to the numbers in Figure 5? If the component parts of the product get bigger, so will the resulting to-do list. Does that look like things are getting tougher? Probably, which certainly makes the case for automation at some level.

So we are left with the original Workload Index and two transformed measures—the normalized Workload Index per worker and an economy-wide to-do list—but which of these is "right"? Which aids and which distorts our understanding of the level of technical risk in cybersecurity? There is more to be done on questions of measurement and cybersecurity [7], but the three measures illuminate three different things and are useful for different purposes.

First, the Workload Index works. It is consistent in how it measures vulnerabilities, providing a replicable time-series measure of the technical problems that plague our systems. The almost u-shaped structure of the Workload Index between 2006 and 2016 suggests, tentatively at least, that technical vulnerabilities might be a bit cyclical. That is useful information to have; firms and the economy can adjust accordingly.

Second, the normalized measure shows that with an expanding IT workforce, the total technical work per worker in the system is not too much worse than it was before. These numbers suggest calm in the face of sensational data breaches that affect millions (or possibly billions). The average network size that people can access once they have breached a system is probably getting bigger, but this measure suggests that keeping any particular part of the system secure on a technical front is not yet a mounting task.

Third, the economy-wide to-do list shows how an increasing worker count and a relatively constant Workload Index can generate a lot of work overall. These numbers suggest that things are getting worse, because the economy is exerting so much effort to keep things afloat. A real trouble here is that more work can mean more room for error, especially if humans remain at the forefront. Additionally, opportunity costs are real. Every hour a worker spends keeping the network safe is an hour which that person could have spent doing something else, something productive rather than protective. At a certain point, the economy-wide to-do list will get too big, the wasted hours will grow too large, and we will have to move towards more automation to keep the networks working and our workers free to do other things.

References

[1] NIST Vulnerability Workload Calculator: nvd.nist.gov/CVSS/v3-calculator.

[2] NIST Vulnerability Workload Index: nvd.nist.gov/Home/Workload-Index.cfm.

[3] Nassim Taleb, *Incerto* tetralogy: *Fooled by Randomness, The Black Swan, The Bed of Procrustes, Antifragile.*

[4] Cybersecurity and higher education: digitalguardian.com/blog/cybersecurity-higher-education-top-cybersecurity-colleges-and-degrees.

[5] Index of Cyber Security: www.cybersecurityindex.org.

[6] US Bureau of Labor Statistics: www.bls.gov/cps/tables.htm#charemp.

[7] E. Jardine, "Garbage In, Garbage Out: Measuring the Effectiveness of Remedial Cybersecurity Policies," working paper.

Writing for *;login:*

We are looking for people with personal experience and expertise who want to share their knowledge by writing. USENIX supports many conferences and workshops, and articles about topics related to any of these subject areas (system administration, programming, SRE, file systems, storage, networking, distributed systems, operating systems, and security) are welcome. We will also publish opinion articles that are relevant to the computer sciences research community, as well as the system adminstrator and SRE communities.

Writing is not easy for most of us. Having your writing rejected, for any reason, is no fun at all. The way to get your articles published in *;login:*, with the least effort on your part and on the part of the staff of *;login:*, is to submit a proposal to login@usenix.org.

PROPOSALS

In the world of publishing, writing a proposal is nothing new. If you plan on writing a book, you need to write one chapter, a proposed table of contents, and the proposal itself and send the package to a book publisher. Writing the entire book first is asking for rejection, unless you are a well-known, popular writer.

;login: proposals are not like paper submission abstracts. We are not asking you to write a draft of the article as the proposal, but instead to describe the article you wish to write. There are some elements that you will want to include in any proposal:

- What's the topic of the article?
- What type of article is it (case study, tutorial, editorial, article based on published paper, etc.)?
- Who is the intended audience (syadmins, programmers, security wonks, network admins, etc.)?
- Why does this article need to be read?
- What, if any, non-text elements (illustrations, code, diagrams, etc.) will be included?
- What is the approximate length of the article?

Start out by answering each of those six questions. In answering the question about length, the limit for articles is about 3,000 words, and we avoid publishing articles longer than six pages. We suggest that you try to keep your article between two and five pages, as this matches the attention span of many people.

The answer to the question about why the article needs to be read is the place to wax enthusiastic. We do not want marketing, but your most eloquent explanation of why this article is important to the readership of *;login:*, which is also the membership of USENIX.

UNACCEPTABLE ARTICLES

;login: will not publish certain articles. These include but are not limited to:

- Previously published articles. A piece that has appeared on your own Web server but has not been posted to USENET or slashdot is not considered to have been published.
- Marketing pieces of any type. We don't accept articles about products. "Marketing" does not include being enthusiastic about a new tool or software that you can download for free, and you are encouraged to write case studies of hardware or software that you helped install and configure, as long as you are not affiliated with or paid by the company you are writing about.
- Personal attacks

FORMAT

The initial reading of your article will be done by people using UNIX systems. Later phases involve Macs, but please send us text/plain formatted documents for the proposal. Send proposals to login@usenix.org.

The final version can be text/plain, text/html, text/markdown, LaTex, or Microsoft Word/Libre Office. Illustrations should be EPS if possible. Vector formats (TIFF, PNG, or JPG) are also acceptable, and should be a minimum of 1,200 pixels wide.

DEADLINES

For our publishing deadlines, including the time you can expect to be asked to read proofs of your article, see the online schedule at www.usenix.org/publications/login/publication_schedule.

COPYRIGHT

You own the copyright to your work and grant USENIX first publication rights. USENIX owns the copyright on the collection that is each issue of *;login:*. You have control over who may reprint your text; financial negotiations are a private matter between you and any reprinter.