For Good Measure
Security Measurement in the Present Tense

DAN GEER

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

Writing a column sometimes requires an "at the time of writing" disclaimer if the situation being described is fluid, de novo, or both. So it is now, which is to say early June, 2020.

By a fluid and de novo situation, I mean the global pandemic known as COVID-19, which is a different beast depending on where you are and how you live. The view from my kitchen table includes a formerly tight lockdown looking soon to be relaxed, pervasive work-at-home for people in technology jobs, a burst of demand for supply chain data, debt burdens too substantial to handle gracefully now or later, and so forth and so on. What might we imagine and, in turn, want to measure under the general topic of "cybersecurity metrics" given the situation? In so many words, here, as in so very much of life, the hard part is getting the questions right—right in the sense of right-for-the-time and supportive of wise decisions. Good questions yield useful answers.

The attack surface comes to mind. I suspect that a material and quite measurable enlargement of the enterprise attack surface due to work at home is hardly a hypothesis. The two components of that expansion that come to my mind as most subtle are routing and sync. How might we measure that expansion, and are estimations on routing and sync the way to go (modeling complex pathways for attack in either case)? Is there a constant of proportionality here and, if so, what might it be? Can its nature be determined by measurement or can measurement merely confirm the assertion of attack surface expansion? Does the fraction of the enterprise's staff working at home reveal a kind of dose-response relationship (a curve of proportionality that demonstrates causality)?

Secondarily, should we expect the changes in the attack surface to show hysteresis? In hysteresis, the output of a system depends not only on its input, but also on its history of past inputs. Put differently, when the force of deformation is relaxed, does the surface spring back to its original position or is the deformation inelastic? Twitter's "work at home forever" comes to mind, but I am thinking more of software installs and changes to standard operating procedure, such as for meetings—installs and changes that won't be de-installed when a COVID-19 vaccine appears. This hysteresis would seem particularly acute in the cybersecurity arena since, as has long been observable, when security products are eclipsed, whether by new organization charts or by new products, existing security products are never de-installed.

The probability of small supplier business failure is surely up. This would imply that the fraction of unmaintained software has risen or will soon rise. Is that measurable? Does the idea of receivership for abandoned software products need measurement (and what kind), or is this just a matter of governmental will? Would measurement help buck up such government will [1]? We'll have this running-but-unsupported situation soon enough once self-modifying code gains autonomy.

What is an "essential" activity and what is not essential is proving to be variously contentious. There's an interesting measure in that, for sure—what fraction of the economy is essential? Rank ordering countries by what fraction of their economy is essential is, likewise, interesting, as would rank ordering cybersecurity functions by whether their operators are deemed essential accordingly. In 2008, we learned a lot about essentialness in and around finance, resulting in an entirely new set of (US) rules for entities that are "too big to fail," or, to be more precise, entities that are SIFIs—systemically important financial institutions. Legalistically, a financial institution is a SIFI if it would pose a serious risk to the economy as a whole were it to collapse. Don't we need that concept in cybersecurity by analogy? Don't we need formal stress testing for computing entities that are not too big to fail but too interconnected to fail [2]? Doesn't cybersecurity eventually, if not now, require such formality? Might we not start now thinking this through?

To the extent that organized opposition (attackers) have an interest in stockpiling 0wned machines, can we measure any uptick in stockpiling in a way that demonstrates causality related to the lockdown crisis? This might be closely related to the routing aspect of attack surface expansion, for example. Or is there a measure that says unequivocally that 0wning a leaf node is nowadays so easy that stockpiling is fiscally irresponsible from the point of view of organized attackers operating as a straight-up business? Or is ransomware now the mechanism of 0wnership? Reported trends in ransomware beg for data on whether the opposition is getting better or the playing field easier to manipulate.

If permanent contraction of enterprises, whether profit or nonprofit, is inevitable, should the cybersecurity workforce enjoy some degree of protection from that contraction? Is there a measure, such as percentage of workforce or percentage of budget, that should be held constant as enterprises contract? (Or, if not that, held above some floor?) Is the skill set among cybersecurity workers a national resource, and do we have numbers to prove it? Is it far-fetched to compare pen-testers adrift in a cybersecurity job collapse to nuclear scientists adrift in the collapse of the USSR? Whatever we've been measuring needs to be re-measured so that trendlines can be established [3].

What about those individuals who were about to enter the cybersecurity workforce, such as recent graduates or those about to be discharged from relevant military service? Do we ensure they find work, or do we have a measure that proves they are unneeded? Are we understaffed with respect to the cybersecurity challenge, overstaffed with respect to the economically provable benefit of cybersecurity practitioners, both, or neither? In many industries, re-opening seems likely to involve replacing humans with algorithms, an ongoing process surely accelerated by the pandemic. Should that be the case in cybersecurity and, regardless of your answer, what might we be measuring here?

In public health, one of the great measurement innovations was the introduction of "quality-adjusted life years" (QALY) as an outcome measure for public health interventions. Do we have some sort of parallel to the QALY measure in cybersecurity? What would it take to have such a measure be defensible as a policy driver? Who should get to set the "adjustment" schedule itself? Also in public health, analyses are often calibrated not just by quality-adjusted life years but also by disability-adjusted life years (DALY, as in disability averted). Is something like DALY more like what we should be measuring in cybersecurity? Or is measurement of either the QALY and DALY sorts built on assumptions that don't actually obtain in cybersecurity? For that matter, where are the tails of distributions getting heavier—the prodromes of black swan events?

In military affairs and emergency management alike, it is all but mandatory that for any given operation or event there be a thorough and dispassionate "after action report" (AAR). Where these are done under a unified command structure such as the Federal Emergency Management Agency [4] or the Department of Homeland Security, their form and scope is itself set by policy. The spirit of the AAR exercise is that of learning lessons from what might realistically be called natural experiments, and formal, fixed output can help make up for the undesigned-ness of any natural experiment. All of which leads us to the question of what should we do in cybersecurity for measuring (and documenting) our version of natural experiments? I would argue that down this path is where we find such things as responsible disclosure mechanisms, bug bounty programs, purposefully opaque software updates, the intermittent appearance of truly novel attacks, and various research results on malware dwell times. Yet to the point here, with lots of cybersecurity AARs to be written in and for the age of pandemics, should we not be measuring and, if so, measuring what, exactly?

One can straightforwardly analogize the "lockdown" strategy as that of decreasing the societal and/or viral attack surface by fiat. I cannot recall as vigorous a purposive reduction in attack surface as the one we saw with COVID-19 (and may, of course, see again should recurrence pick up). On the biologic side, the lockdown was supported by rather an explosion of creative modeling. Take just the one example of wearing a face mask; it protects others from your spew more than it protects you from others'. The benefit of wearing a mask is not transitive, but the risk of not wearing one is (transitive). That's a bit like not allowing your computers to be part of a botnet; it doesn't protect you from others but rather it protects others from you. We need a measure for how much your computing is a danger to others [5], though, of course, such a measure (and the policy it would support) is likely to be met with the same mix of hostile compliance that mandatory face masks exhibits. What should we measure? What should we model? How might we think quantitatively on what sort of cyber pandemic would require turning off electronic commerce until a suite of not yet designed patches (vaccines) could be rolled out to machines young and old alike? Are those countries experimenting with disconnecting from the public Internet [6] on to something measurable?

Health policy and management is perfectly happy (and for good reason) with herd immunity; should we be [7]? What if the exposed fraction of Internet users is largely concentrated in one jurisdiction or among one class of users? Or, as described in the prior reference, how we measure would be correlated with what we conclude is our societal mandate—would we prefer to minimize harm (like reserving scarce vaccine for the young and the old) or would we prefer to minimize transmission (like reserving scarce vaccine for health care workers and undertakers)? Don't answer "both."

In Summary

What I am trying to get at is that what actions we take, at least what considered actions we take, is as influenced by what we measure as it is by what those measurements show. Thinking it out in advance sure beats decision-making under the influence of adrenaline.

I close with a quote from John Foster Dulles, Secretary of State under President Eisenhower:

> The measure of success is not whether you have a tough problem to deal with, but whether it is the same problem you had last year.

References

[1] D. Geer, "On Abandonment": geer.tinho.net/ieee/ieee.sp.geer.1307.pdf.

[2] D. Geer, "For Good Measure: Stress Analysis," *;login:*, vol. 39, no. 6 (December 2014): https://www.usenix.org/system/files/login/articles/login_dec14_13_geer.pdf.

[3] D. Geer, "For Good Measure: Cyberjobsecurity," *;login:*, vol. 45, no. 1 (Spring 2020): https://www.usenix.org/system/files/login/articles/spring20_14_geer.pdf.

[4] For an example after action report, see emilms.fema.gov/IS130a/groups/57.html.

[5] D. Geer, "CyberGreen Metrics," October 2016: www.cybergreen.net/img/medialibrary/CyberGreen%20Metrics%20v.2.pdf.

[6] Russia, December 2019, for example.

[7] D. Larremore, D. Geer, "Progress Is Infectious," *IEEE Security & Privacy*, vol. 10, no. 6 (November 2012): geer.tinho.net/fgm/fgm.geer.1211.pdf.