# For Good Measure
## When Opinion Is Data

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

Obvious to all, the sea of data is rising. It's a remarkable thing really. Even if all you can remember is 10 years back, the comparison of "then" with "now" is pretty startling. No, that does not qualify as news, but to reparse Orwell's "Who controls the past, controls the future: who controls the present, controls the past," the data "we" collect now is what will soon enough become the past for a data-driven world. If that data past comes to exert a force in some sort of proportion to its volume, is there, or will there be, any room for mere human opinion?

Cybersecurity has long had a measurement problem. Progress has certainly been made, both in the pages of this publication and elsewhere. Defenses now include mass data collection and tools whose main job is to reduce data volume to something that is straightforwardly actionable. In the Orwell sense, the algorithms that collect and reduce the instrumentation data are coming to control if not the present itself then our understanding of the present. In due course, the "actionable" becomes the automatically acted upon, that is to say that algorithms are trusted to do what we seem unable to do—to protect us from other algorithms. Such is progress.

Yet the nuance here is that the algorithms are, by and large, uninterrogatable—they cannot be meaningfully asked why they made such and such a decision. The outcome of action, not the reason for action, becomes the only check and balance that we humans have at all. This may be a tradeoff that is not just inevitable but welcome, welcome in the sense of freeing front-line cybersecurity staff from having to juggle a million balls all at once. At the same time, if you/we cannot examine the reasoning behind an automatic action but only react to the outcome of it, what then do we know about the present? What kind of past will the accumulating data create? Behaviorally oriented cybersecurity is entirely crafted along these lines, the line of learning enough about the recent past to be able to tell that the present is diverging from that past and, ipso facto, algorithmically control the future. What then is the role of the human in the loop?

The Index of Cyber Security (ICS) was created six years and a little more ago on the premise that we didn't know enough about the details of cybersecurity to make prediction and planning really possible—that "the present" was (is) a bit of a miasma and, as such, the best and only trustable prediction of the future was to be found in the pooled opinions of front-line cybersecurity practitioners. As with the oft-noted "wisdom of crowds," ours was not a search for the single smartest oracle but rather a pooling of opinion from a body of experts whose views were tempered by the heat of daily practice. Speaking for myself and my colleague in this project, we think that the need for pooled expert opinion is greater than ever, both between practitioners (as with the ICS) and inside each firm that is itself large or connected enough to be a constant target.

A developed muscle that is not exercised will atrophy. A developed skill that is not exercised will atrophy. If we humans are to remain the ultimate decision makers regarding our fate,

then our ability to form strong opinions must not be left unexercised, it must not be left to atrophy. The desire for automaticity runs toward setting the stage for an atrophy of some skills; choosing what to let go may require the greatest of wisdom. Perhaps, then, the state of our wisdom is worth close attention. To illustrate that point, consider this ICS question:

Your organization is likely more reliant on the cloud than you think. According to Symantec's Internet Security Threat Report, the average enterprise organization was using 928 cloud apps, up from 841 earlier in the year. However, most CIOs think their organization only uses around 30 or 40 cloud apps. Reliance on the cloud goes beyond the traditional infrastructure hosting arrangement. Unknown to IT, the "business" will often sign up for SAAS services on the cloud where data (or metadata at least) gets out on the cloud.

What is your assessment of your security organization's handle on cloud engagement:

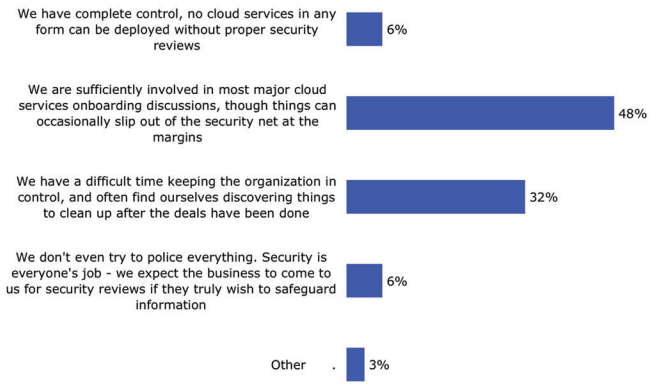


**Figure 1**

That question above and its answers by a pool of front-line cybersecurity people is illustrative—both of the spread of opinion and its logic. That we can ask practitioners such a question is the interrogatability part. That some entities centralize control while others delegate responsibility is no real surprise but is still worth noting insofar as it says pretty clearly that no single "right" answer has come along.

Let's try another:

After years of study, we still do not seem to be able to agree on the question of vulnerabilities and, in particular, matters of their discovery, use, retention, and disclosure. Policy constraints vary across countries like night and day. These are strategic issues or, should we say, Strategic Issues that fully prove that cybersecurity and the future of humanity are conjoined now. Allowing for ambiguity, which of these directions should free-world governments favor:

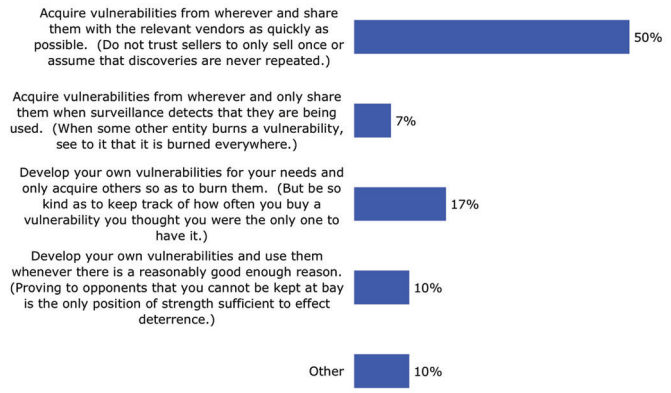


**Figure 2**

As with the first example, the spread of opinion is valuable in and of itself. Does not the preponderance of the first option, to acquire vulnerabilities from wherever and share them with the relevant vendors post-haste reflect a strong prediction on the part of the respondents about what they expect the vulnerability situation to be in future? Would an algorithm fed by a sensor network come to the same conclusion?

Let's try a third:

Newly discovered vulnerabilities create workload for defenders that is immediate—in the form of security updates and patches to apply—and workload that is deferred—as everything built and deployed from that point on has to be inoculated against the continuously accretive database of known weaknesses. Yet this work cannot be perfectly sufficient, as Mirai has shown; the capabilities of the attackers can increase even if the defense is doing everything right for their organization.

How have you been seeing your workload fluctuate over the past year:

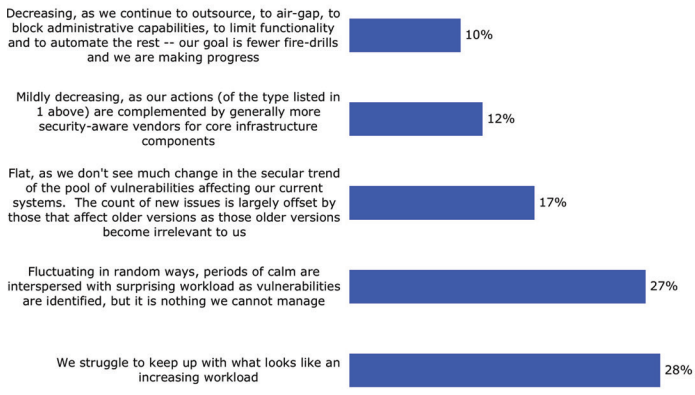


**Figure 3**

Here, the respondents' opinions are certainly predictive about the future of their own practice, and, from that, one can make broader statements about the cybersecurity situation in general. This human judgment seems better than any sensor network-driven machine learning could be expected to deliver. Of course, sometimes it is not a question of data but rather of the handling of data, such as this fourth example:

Information sharing with the government, even after large incidents, is an activity fraught with anxiety and stress. Differential reporting by the victim targets means the data that public authorities have is not useful for rational planning. Some target entities will report; some will not. Has the time come to have an escalation rule for sharing of information about attacks?

We do this with different rationales in some contexts, such as when we require prompt and detailed attack information from defense contractors to Pentagon authorities, when state laws force disclosure if a customer's credit card or other personal information is exposed, and when the SEC requires the announcement of security breaches that materially impair corporate operations. Has the time come for a mandatory reporting regime for all events that are above some threshold of severity?

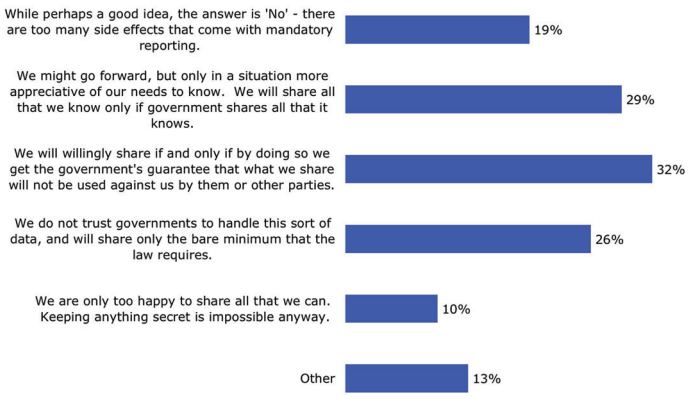


**Figure 4**

Collectively, these questions illustrate what shared, expert opinion can mean, and it seems unlikely that algorithms would take over these areas of informed choice, but 10 years ago we would not have guessed what algorithms have taken over today either. While we can (and will) ask the ICS respondents about the role of automation in the near-term future, our imagination may not be up to the task of asking the right questions. By all means, make suggestions as to what questions we should ask. If you are, yourself, a front-line security practitioner, then please consider becoming one of our respondents (it will cost you 10 minutes a month, and you will see a lot of analysis that we reserve for our respondents—though we'll happily provide a sample to help you make a decision).

Nevertheless, at the end of the day, the biggest question is whether a human in the loop is a failsafe or a liability. We favor the "failsafe" view, but to keep and maintain that a human in the loop is a failsafe, they have to actually be in the loop. Being an observer of algorithms that don't ask (permission) and don't tell (what it is they are doing) won't keep the practitioner in fighting trim. There's no such thing as a free lunch…