For Good Measure

Numbers Are Where You Find Them

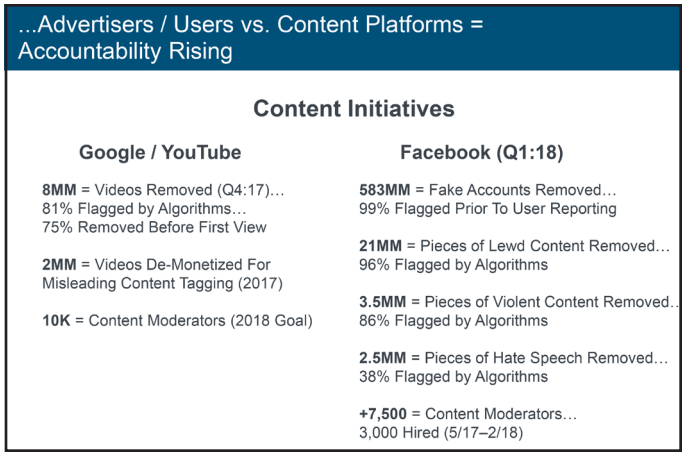
DAN GEER

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

It will come as no surprise that 50, or even 20, years ago inquisitive minds were often at a loss for bodies of numbers upon which they could rely. Putting aside the precise meaning of "rely" for the moment, a shortage of numbers is less and less a reason for inaction in any domain. Just as obvious as the sunrise, soon enough the issue will be too many numbers. Sensors, radios, and AI, oh my.

Security metrics study is possibly out in front of some other fields, but only so much and likely not even that for much longer. The idea that managing a risk requires measuring that risk or its precursors has long since become standard operational thinking in the security game, yet we are living proof that while collecting numbers is necessary, it is not sufficient to deliver security.

Some would argue that it isn't our tools and our scorekeeping (with numbers) that is the "thing" that is not sufficient—rather, it is incentives that are wrong. Whole conferences are on this topic, and there is no way to summarize them in the context of this column, but study of, and suggestions for, incentive structures, be they rewarding or punishing, are surely needful. As an example, the organizers of the Code Conference (CodeCon) said, "[For] 2018, we invited the people in charge of enforcing regulations, and those creating new ones." That's a stab at incentive structures to be sure, but let's specifically look at some numbers from Mary Meeker's "Internet Trends 2018" slide deck at CodeCon, beginning with Slide 99 [1].

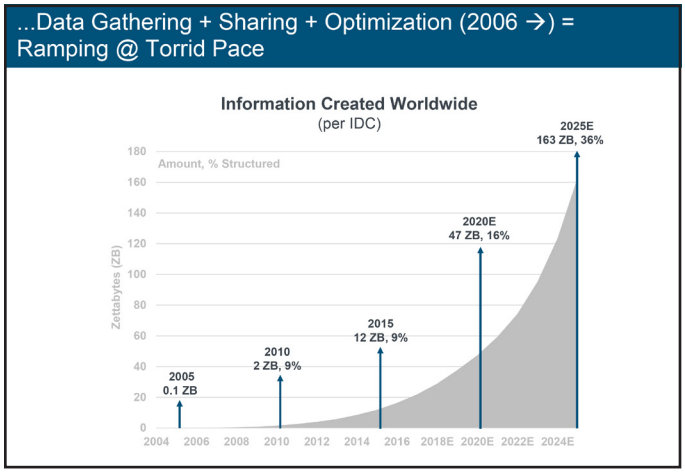


Content initiatives, Slide 99

Note the role of algorithms in the above, which, for the purpose of this column, we will take as a form of security metrics even if the algorithms in question are not open for inspection. Algorithms as censors is a worthy topic in its own right.

With Google/YouTube, that 81% were flagged by algorithms presumably means that the average Content Moderator sees those algorithms as automated assistance to making faster

Online prices are falling, Slide 111



Projection of global information creation, Slide 189

decisions. A month before Meeker's speech, *The Guardian* [2] said this about the algorithms, per se:

> Those systems broadly work in one of three ways: some use an algorithm to fingerprint inappropriate footage and then match it to future uploads; others track suspicious patterns of uploads, which is particularly useful for spam detection. A third set of systems use the company's machine learning technology to identify videos that breach guidelines based on their similarity to previous videos. The machine learning system used to identify violent extremist content, for instance, was trained on 2 million hand-reviewed videos.
>
> While machine learning catches many videos, YouTube still lets individuals flag videos. Members of the public can mark any video as breaching community guidelines. There is also a group of individuals and 150 organisations who are "trusted flaggers"—experts in various areas of contested content who are given special tools to highlight problematic videos. Regular users flag 95% of the videos that aren't caught by the automatic detection, while trusted flaggers provide the other 5%. But the success rates are reversed, with reports from trusted flaggers leading to 14% of the removals on the site, and regular users just 5%.

Facebook's use of algorithms is undoubtedly similar to that of Google/YouTube.

But measurement is not a problem just for us here in security metrics land; take something as important as economics. Everyone knows something about the Consumer Price Index (CPI). As Wikipedia puts it, "In most countries, the CPI, along with the population census, is one of the most closely watched national economic statistics." Yet even the calculation of the CPI is having trouble these days, as Slide 111 shows.

Think of the spread of things that the CPI is baked into, from labor contracts to entitlements to financial instruments to you-name-it. Surely the CPI is easier to measure than security, but here we are.

Of course, everyone knows that the world is creating lots of data. Defining "structured" data as "data that has been organized so that it is easily searchable and includes metadata and machine-to-machine (M2M) data," we have (by way of the market intelligence firm IDC) the curve you see in Slide 189.

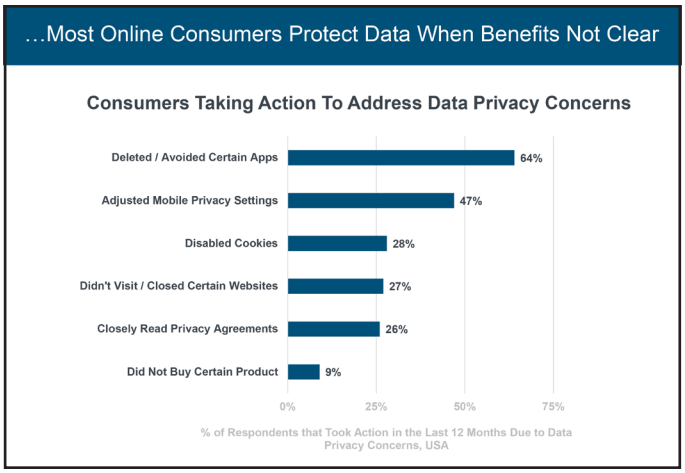
For those of us working in data protection, the message is obvious—data protection must be automated; the algorithms have to make the "kill decisions." And other algorithms will have to summarize things for us, summaries that will be ever more distant from the raw numbers.

Putting aside the argument over whether security and privacy are mutually supportive or fundamentally at odds, Slide 206 has a few somewhat encouraging numbers that consumers are at least thinking about it:

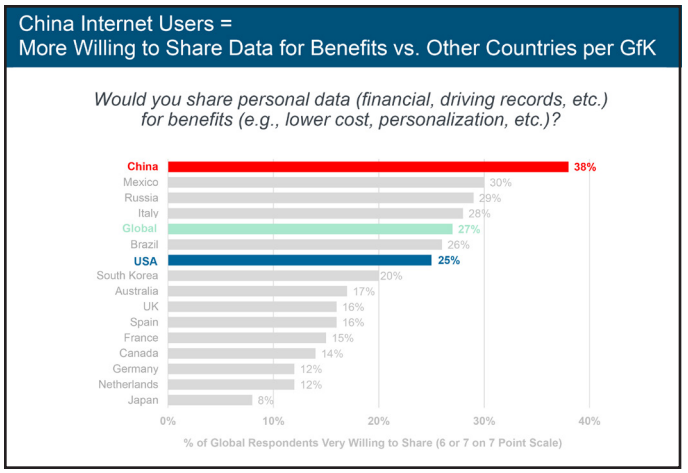
On the other hand, trading short-term gain for long-term risk is still blithely popular, but, as measured by the German marketing firm GfK, blitheness is culturally diverse—see Slide 223.

That one slide, Slide 223, probably says more than we know how to evaluate both as to privacy (the question GfK actually investigated) and to security (as in risk/benefit tradeoffs generally). Later on (Slide 266), the founder of Slack hits the nail on the head: "When you want something really bad, you will put up with a lot of flaws." We, the global "we," want our toys ever harder, ever faster. There've been a lot of demonstrations of that phenomenon, but let's use *The Economist*'s numbers in Table 1 [3].

To get adoption rates accelerating like that a lot of flaws must be put up with, security flaws in particular, one might presume,



How consumers address data privacy concerns, Slide 206



Willingness to share data by country, Slide 223

since security is so generally a feature worth adding only after there is good consumer uptake.

So where does this get us? It is not as if anyone needs to be told that things are changing faster than we understand. It is not as if the present author's selected numbers from the "Internet Trends Report" are unbelievable individually, but collectively they predict a world where prediction (as we humans understand the term) is less and less possible because of the number of moving parts, their opacity, their interdependence, their cardinality, their specificity of purpose, their autonomy, their speed. Clausewitz would no doubt call this a deepening fog of war. Modern military doctrine trades off precision and certainty for speed and agility, or, as Army Chief of Staff Gen. Mark Milley says [4], "On the future battlefield, if you stay in one place longer than two or three hours, you will be dead." Is that not the future of cybersecurity in a nutshell?

Years until used by one-quarter of American population	
46	Electricity
35	Telephone
31	Radio
26	Television
16	Personal Computer
13	Mobile Phone
7	The Web

Table 1: Technology adoption

References

[1] M. Meeker, "Internet Trends 2018": http://www.kpcb.com/file/2018-internet-trends-report.

[2] A. Hern, "YouTube Reveals It Removed 8.3m Videos from Site in Three Months," *The Guardian*, April 23, 2018: https://www.theguardian.com/technology/2018/apr/24/youtube-reveals-it-removed-83m-videos-from-site-in-three-months.

[3] "Happy Birthday World Wide Web," *The Economist*, March 12, 2014: https://www.economist.com/graphic-detail/2014/03/12/happy-birthday-world-wide-web.

[4] S. Freedberg, Jr., "Miserable, Disobedient and Victorious: Gen. Milley's Future US Soldier," *Breaking Defense*, October 5, 2016: https://breakingdefense.com/2016/10/miserable-disobedient-victorious-gen-milleys-future-us-soldier/.