For Good Measure
Implications of the IoT

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

Everybody is predicting great things, within varying interpretations of the word "great," for the Internet of Things. You are doubtless tired of hearing that. As good an answer as any to the question "When was the IoT born?" is when the number of connected devices exceeded the number of humans, while as good an answer as any to the question "What is a thing?" is any data-handling entity that cannot be found in contempt of court.

To remind ourselves of the basic numbers, Figure 1 is IoT size, Figure 2 is how many humans, and Figure 3 is thus the number of things per person.

The counts of human population are probably pretty close to correct. The counts for the IoT are surely arguable. The smooth curve in Figure 1 is simply Cisco's calculated exponential from their 1992 figure of one million devices to their estimate of 50.1 billion in 2020. In each of Figures 1, 2, and 3, the values for 2015 and for 2020 are highlighted. Last year might well have been the real birthdate of the IoT, in other words. Or maybe you don't want to compare all humans to the number of connected devices, only humans who are connected. Globally for 2015, 46% of humans were connected, and the year when there were more *connected* devices than *connected* humans was accordingly earlier, perhaps 2010. Regardless, the IoT is between infant and toddler.

There are two aspects to oncoming growth like this that are directly relevant to public policy, one germane to "For Good Measure" and one not. The "not" is the lifetime resource cost, including energy cost of manufacture, operation, and disposal, that an exponentially increasing number of powered devices necessarily represents. Regulators are looking at this with whatever passes for glee in such places—IT, broadly defined, already accounts for 5–10% of the developed world's energy use.

The other (and germane) aspect is that of attack surface. We obviously don't know what the attack surface of the IoT is—we can scarcely imagine what "attack surface" means in context or even if it means something unitary and evaluable, but whatever that attack surface is, given (genuinely) exponential growth in counts of devices, it is hard to imagine that there is no risk being added to the connected parts of the globe. Just to keep aggregate risk static requires that the risk per device not only fall faster than the curve of deployment rises, but faster still if it is to drown out the legacy risk of devices previously installed. That is a tall order. Ergo, we should doubtless assume for planning purposes that we will see a significant, ongoing increase in the aggregate attack surface.

Or not. Does the attack surface construct even make sense when we are contemplating 10^{10} devices? Clearly, redundancy can contribute real survivability value for a sensor deployment—one broken sensor just doesn't matter if there are others doing the same data capture. One is reminded that network layout can deliver resistance to random faults or resistance to targeted faults but not both. Could not the same be said of sensor data and its roll-up to decision support, that one's threat model has to make some tradeoffs between being invulnerable to random sensor failure and being invulnerable to targeted (intentional) sensor failure?

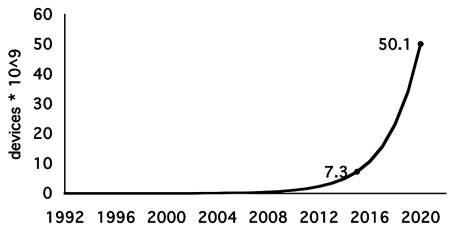

Figure 1: Billions of devices

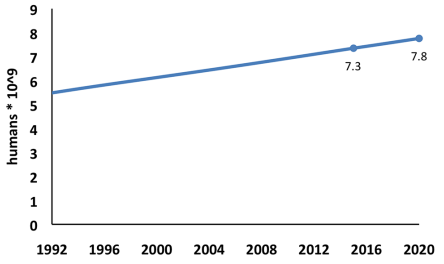

Figure 2: Billions of humans

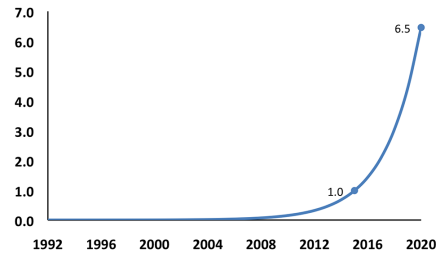

Figure 3: Devices per human

That said, the real question is what should we measure either for the numerator (risk) or the denominator (normalization to something)? *Hourly* data traffic today exceeds annual data traffic of only 10 years ago, and IoT devices are nothing if not traffickers in information. So is data volume the base proportionality constant for an "IoT attack surface"? Or is the attack surface proportional to the percentage of small devices that have Turing-complete remote management interfaces? Or is the attack surface proportional to the minimum practical latency between problem discovery and effective problem repair (thinking now of 1000 million devices with a common mode vulnerability just discovered).

As you know, there is much focus today in the security product market on behavioral security, on accumulating an idea of what routine operation looks like the better to detect badness early, but is anyone actually proposing watching 10^{10} data sources for actionable anomalies? Presumably not, but does that tell us of a latent need or does it tell us something else again? Something about redundancy or about minimizing dependence on singleton devices? Something about trading off the risks of single points of failure against the risks of common mode failure?

It is likely that there are no best answers, and that all answers will be context dependent—a threat model to rationalize attack surface measurement for, say, medical care will be something entirely different than for, say, shopping mall inventory control. Any context that actually matters will have to have an attack surface metric (or something like it) that scales well; Qualcomm's Swarm Lab at UC Berkeley has notably predicted 1000 radios per human by 2025, while Pete Diamandis' *Abundance* calls for $45x10^{12}$ networked sensors by 2035. These kinds of scale cannot be supervised, they can only be deployed and left to free-run. If any of this free-running is self-modifying, the concept of attack surface is just plain over. So, too, is the concept of trustworthy computing, at least as presently understood.

In any case, we are past the point of no return here. The IoT and its scale make most of our gross measures (like attack surface, say) into historical curiosities. The present author has long thought that the pinnacle goal of security engineering to be "No Silent Failure," and with the IoT at its predicted scale, perhaps that goal will now meet its most formidable challenge. It may be that for the IoT we security metricians will have to start over. It may be that our metrics for the IoT will be less observational and more analytic, such as "How much silent failure is tolerable?" Surely adding 10^{10} devices to the connected world increases its complexity, and more complexity means less system predictability, which conflicts with security goals. Distributions of events that we can detect and count today are looking more and more like power laws. If that is an emergent truth and not our confusion, then it is Nassim Taleb's prediction that matters most: "[We are] undergoing a switch between [continuous low grade volatility] to...the process moving by jumps, with less and less variations outside of jumps."

Yes, predictions around the IoT are a dime-a-dozen, many of them are non-falsifiable, and no forecaster ever got fired for adding an extra zero to some rosy daydream. What's your wager?