LAWFARE

HARD NATIONAL SECURITY CHOICES

Where the Science is Taking Us in Cybersecurity

By Dan Geer

Friday, May 29, 2015 at 10:37 AM

Science tends to take us places where policy cannot follow. Policy tends to take us places where science cannot follow. Yet neither science nor policy can be unmindful of the other. Here I will confine myself to six points where I see science, including applied science, asking us to look ahead (The following is necessarily short; for a longer treatment of the science of security, per se, see "T.S. Kuhn Revisited," keynote to biennial meeting of NSF Principal Investigators, February 6, 2015.):

1. Identity
2. Ownership as perimeter
3. Control diffusion
4. Communications provenance
5. Everything is unique
6. Opaqueness is forever

1. Identity

Miniaturization will continue its long-running progression and, in consequence, devices will continue to proliferate into spaces in which they were never before present. Burgeoning proliferation demands device autonomy, and will get it. For autonomy to not itself be a source of irredeemable failure modes, devices will have individual identities and some degree of decision-making capacity.

As device counts grow, device identity eclipses (human) user identity because user identity can be derived from device identity insofar as the proliferation of devices means that users are each and severally surrounded by multiple devices, devices whose identity is baked into their individual hardware, as is already the case in mobile telephony.

There is then neither need nor process to assert "My name is Dan" as Dan's several devices will collectively confirm that this is Dan, perhaps in consultation with each other. As per Zuboff's Laws (Everything that can be automated will be automated. Everything that can be informated will be informated. Every digital application that can be used for surveillance and control will be used for surveillance and control.), all devices are therefore sensors and as the devices themselves have immutable device identities, Dan's claim to being Dan is decided algorithmically. And distally.

Cryptographic keys for users thus become irrelevant as devices will have them, thereby freeing users from key management, much less password drills. The

Fifth Amendment is entirely mooted, as Courts have already ruled that only something you know is protected thereunder, not something you are or have—that is to say that production of devices under subpoena cannot be thwarted. (See *Virginia v. Baust*, CR 14-1439, 28 Oct 2014.)

The longstanding debate over whether identity should be name-centric (where "Dan" is the identity and some key is an attribute of "Dan") or key-centric (where the key is the identity and "Dan" is an attribute of that key) is thus decided in favor of key-centricity though the keys are now held in a fog of small devices. This setting mimics how a stratum of elite people carry neither identification nor money—in the context of their retinue there is no need for such.

For the result of this data fusion to not be a unitary identity for the individual user, policy will have to demarcate data fusion with a vigor it has never before dared. (Privacy is the effective capacity to misrepresent yourself, ergo, it is your devices that give pawns to fortune. See "Tradeoffs in Cyber Security.")

2. Ownership as Perimeter

The paradigm of cybersecurity has long been perimeter control, but that same proliferation of devices rewrites the calculus of what is a perimeter. It is clear that the design of the Internet as we now know it rests on two principles above all others, preferential attachment (See Barabasi L & Albert R, "Emergence of scaling in random networks," Science, v286 p509-512, October 1999) and end-to-end communication protection. Preferential attachment yields scale-free network growth that, in turn, maximizes network resistance to random faults; Internet build-out could *not* have happened otherwise. The end-to-end principle is and has been the fuel for innovation—as end-to-end scales, whereas permission seeking does not.

Both of those principles are under stress. First, the S-curve of Internet growth passed its inflection point in November of 2008, at least for hosts addressable by name, and since that time growth rates have slowed. (This measurement tool is insensitive to the Dark Net which heavily overlaps the Internet of Things.)

Second, random faults no longer comprise the availability risk they once did, all the while carriers and governments alike clearly want non-preferential attachment, carriers in their desire for economic hegemony, free-world governments in their desire for attribution, and unfree-world governments in their desire to manipulate information flow.

Add in the proliferation of small devices and the paradigm of cyber security can no longer be perimeter control. To take but one example, let's count cores in the Qualcomm Snapdragon 801. The central CPU is 4 Cores, the Adreno 330 GPU another 4, Video Out is 1 more, the Hexagon QDSP is 3, the Modem is at least 2 and most likely 4, Bluetooth is another 1 as is the USB controller and the GPS. The Wifi is at least 1 and most likely 2, and none of this includes charging, power, or display. That makes somewhere between 18 and 21 cores. In the vocabulary of the Internet of Things, I ask you whether that is one thing or the better part of two dozen things? It is pretty certain that each of those cores can reach the others, so is the perimeter to be defended the physical artifact in the user's pocket or is it the execution space of each of those cores?

I looked at seven different estimates of the growth of the Internet of Things as a market phenomenon, everything from smart electric meters to networked light bulbs to luxury automobiles, and the median is a compound annual growth rate of 35 percent. If perimeter control is to remain the paradigm of cybersecurity, then the number of perimeters to defend in the Internet of Things is doubling every 17 months.

So what is to be the perimeter of control from a cybersecurity point of view? Is it ownership that demarcates perimeter? More and more of user capability is controlled by licensure, not ownership in the dictionary sense of the word "ownership." The science is taking us away from ownership conferring cradle-tograve control towards a spectrum of temporally constrained permission granting; I can give you my bed, but I cannot give you my iTunes. Self-driving cars are as good an illustration as any; over-the-air auto-update of firmware will not be optional in either time or place and vehicle-to-vehicle communication will do route selection in the name of the common good. In a digital world, nothing comes without strings attached.

3. Control diffusion

As has been shown in finance, if one entity can do high-speed trading then all must, but whereas predatory and/or unstable trading is subject to a quantum of regulatory control, cyber predation is not, and cyber predators have zero legacy drag. As such, turning over our protections to machines is inevitable. Science and startups alike are delivering a welter of automation for protection, most not involving recondite algorithms but rather big-data fueled learning about what is normal, the better to identify that which is not normal and thus suspect.

I leave to any policy discussion the question of whether the speeds at which cyber security automation must run will even allow occasional interruption to ask some human operator for permissions to act, or must cyber kill decisions be automated on the argument that only when so automated can they respond in time. If the latter holds, and I am certain that it will, science will be under the gun to encode human ethics into algorithms that will free run. Put differently, I predict that it is in cyber security, per se, where the argument over artificial intelligence will find its foremost concretization. Frankly, I side with Hawking, Gates, and Musk on such matters. As an example of an unevaluable vignette, the self-driving car will choose between killing its solo passenger or fifteen people on the sidewalk. Many's the example of airplane pilots sacrificing themselves to avoid crash landing in populated zones.

Coupled with algorithmic user identification, control thus enters a state where trust is multi-way, not one-to-one. It is hard to overestimate just how much the client has become the server's server. Take Javascript, which is to say server-side demands that clients run programs as a condition of use, or web screens recursively assembled from unidentifiable third parties; the HTTP Archive says that the average web page now makes out-references to 16 different domains as well as making 17 Javascript requests per page, and the Javascript byte count is five times the HTML byte count. A lot of that Javascript is about analytics, which is to say surveillance of the user.

But as a practical matter, any important control needs an override, such as for medical emergencies. Barring national security situations, such override is closer to a failure, a failure that must not be silent. If the pinnacle goal of security engineering is "No silent failure," then the as-yet-unmet challenge is how to design cybersecurity such that it never fails silently. There is scientific work to be done here—full automation of cyber security maximizes the downside cost of falsely positive indicators of attack.

4. Communications Provenance

Provenance of network traffic will rise to new importance unrelated to quality of service or transport neutrality.

Delegation of credentials has heretofore been driven by executives delegating correspondence handling to their assistants; as devices proliferate, delegation of credentials and authority becomes a necessity across the board, at least for First World digerati. Take loading a web page in a browser: the browser does proxying, nameservice lookup, etc., and eventually loads that page plus subsequent web page dependencies, probably from other sites. In other words, there are various levels of "who" actually requested what, such as what piece of Javascript invoked Google Analytics. As a one-off experiment, I looked at the topmost page of cnn.com; there I found 612 HREFs across 38 hosts in 20 domains even without evaluating the 30-odd Javascripts there. Competent scientists are studying the issue of how to characterize multi-dimensional attack surfaces, and we should attend their results.

Because cyber security is to remain the driving reason for egress filtering, provenance—as in “Who ordered this page?”—is the crucial variable for intelligent flow control. If cyber integrity of the browser platform itself is to remain the topmost user goal, then agency—again as in “Who ordered this page?”—is likewise the most important variable for permission decisions.

This need will be met with traffic analysis extending into the execution environment, which will come as no surprise to this audience. What may be instructive, however, is that when the civilian public came to need encryption, within a decade the commercial sector caught up to the military sector in the application of cryptography. Now the marketeers are driving the commercial sector to catch up to the military sector in traffic analysis. How the traffic analysis that marketeers demand (and will get) meshes with the traffic analysis on end-users delegating human authority to their growing constellation of devices remains to be seen, but with dual demand for traffic analysis, the commercial sector will fill that demand one way or another.

But even if the public and the marketeers want some kind of traffic analysis that is of a toy variety compared to what the military sector needs, there are two other considerations at play. One consideration is that a non-negligible fraction of Internet backbone traffic cannot be identified by protocol—in other words, it has no provenance and is likely peer to peer. While intentionally obscure traffic may as easily be pedophiles as heroic freedom fighters posting unexpurgated calls to arms, in a world where it is the machines that provide the cyber security by learning what is normal so as to tag what is abnormal, the pedophiles and the freedom fighters will stand equal chances of being blocked, if not outed.

The other consideration is junk traffic, meaning traffic whose emitter is on auto-pilot but whose purpose is long defunct. Years ago, my colleagues spent some time trying to figure out what was calling one of our dialup numbers. In the end, it turned out to be an oil tank in an abandoned building that was outfitted to request a fill when needed, and we had inherited the number to which such requests had once gone.

Junk traffic will have to be dealt with via provenance or some discoverable correlate of provenance. Perhaps we will remanufacture spam detection for this purpose. Perhaps traceability will become the rule of law as soon as geolocation applies to the Internet as much as it now applies to cell phone triangulation.

5. Everything is Unique

Science is fast teaching us that everything is unique if examined at close enough detail. Some of it you already know; facial recognition is possible at 500 meters, iris recognition is possible at 50 meters, and heart-beat recognition is possible at 5 meters. Your dog can identify you by smell; so, too, can an electronic dog’s nose. Your cell phone’s accelerometer is plenty sensitive enough to identify you by gait analysis. A photograph can be matched to the camera from which it came as well as a bullet can be matched to the barrel through which it passed. Some apartment building owners now require that tenants provide a DNA sample of their dog so that unscooped poop can be traced.

When everything is detectably unique, decision support of many sorts becomes possible. Assessing nuances, such as whether you are angry, will be embedded in automatons. Accountability will doubtless be extended to ever-more-minor behaviors. That heartbeat recognition technology is already slated to be part of automobiles. Courtroom alibis will soon be backed by cybersecurity-like evidence, noting that because an alibi involves evidence of innocence rather than of guilt, the privilege against self-incrimination is not implicated and is, instead, subject to compelled disclosure. The testimony of spouses against each other will be unnecessary—their devices will do.

6. Opaqueness is Forever

Where data science spreads, a massive increase in tailorability to conditions follows. Even if Moore's Law remains forever valid, there will never be enough computing, hence data driven algorithms must favor efficiency above all else. Yet the more efficient the algorithm, the less interrogatable it is, that is to say that the more optimized the algorithm is, the harder it is to know what the algorithm is really doing.

The more desirable some particular automation is judged to be, the more data it is given. The more data it is given, the more its data utilization efficiency matters. The more its data utilization efficiency matters, the more its algorithms will evolve to opaque operation. Above some threshold of dependence on such an algorithm in practice, there can be no going back. As such, if science wishes to be useful, preserving algorithm interrogatability despite efficiency-seeking, self-driven evolution is the research grade problem now on the table. If science does not pick this up, then Lessig's characterization of code as law is fulfilled.

Implications: Why this Matters

There is no argument whatsoever that the proliferation of devices and information are empowering. It is categorically true, not to mention obvious, that technology is today far more democratically available than it was yesterday and less than it will be tomorrow. 3D printing, the whole "maker" community, DIY biology, micro-drones, search, home automation, constant contact with whomever you choose to be in constant contact with—these are all examples of democratizing technology. This is perhaps our last fundamental tradeoff before the Singularity occurs: Do we, as a society, want the comfort and convenience of increasingly technologic, invisible digital integration enough to pay for those benefits with the liberties that must be given up to be protected from the downsides of that integration? If, as Peter Bernstein said, risk is that more things can happen than will, then what is the ratio of things that can now happen that are good to things that can now happen that are bad? Is the good fraction growing faster than the bad fraction or the other way around? Is there a threshold of interdependence beyond which good or bad overwhelmingly dominate? Now that we need cybersecurity protections to the degree that we do, to whom does the responsibility devolve? If the worst laws are those that are unenforceable, what would we hope our lawmakers say about technologies that are not yet critical but soon will be?

Growth in personal power has meant that heretofore military needs, like traffic analysis, have become common needs. One then asks whether entities under attack, be they enterprises, carriers, or individuals, can garner enough provenance to engage in "strike back" and, if so, should they? As a righteous societal good, we, by law, require that persons with disabilities not be prevented from the fullest possible participation in our society. Can we find the wisdom to do something equivalent for those who cannot or, to the point, will not adopt the plethora of technologies that are redefining what "full participation in society" means? Is preserving such options the surest way to prevent a common mode digital failure?

The need for what we have heretofore called cybersecurity is now so varied that it is no longer a single field but many. There are over 800, perhaps over 1000, cybersecurity startups in some stage of the funding game, a fair fraction of them spinouts from highly focused university research projects. Generalists such as myself cannot be replaced—there is too much for the novitiate to learn. The core knowledge base has reached the point where new recruits can no longer hope to someday become competent generalists, serial specialization is the only broad option available to them. Cybersecurity is perhaps the most difficult intellectual profession on the planet. Ray Kurzweil is beyond all doubt correct; within the career lifetime of nearly everyone in this room, algorithms will be smarter than we are, and they will therefore be called upon to do what we cannot—to protect us from other algorithms, and to ask no permission in so doing. Do we, like Ulysses, lash ourselves to the mast or do we, as the some would say, relax and enjoy the inevitable? What would we have science do?

Daniel E. Geer, Jr., Sc.D., serves as Chief Information Security Officer at In-Q-Tel, the strategic investment partner of the U.S. intelligence community, and has held C-level positions at six startups over the past two decades. Prior to that, he led systems development at MIT's Project Athena out of which came many of the underpinnings of today's Internet and, earlier still, worked in medical computing within Harvard's various teaching hospitals. He provides advice and counsel to numerous Federal agencies, and has been before Congress five times. Dr. Geer's degrees are in Biostatistics from the Harvard School of Public Health and in Electrical Engineering from MIT, and he has been honored with the Lifetime Achievement Award of the USENIX Association.

0 3 53 0

Add a comment...

Comment

Facebook social plugin

Related Posts:

1. **The Full Glare of European Hypocrisy on Surveillance**
2. **@War: The Rise of the Military-Internet Complex**
3. **An Interview with FBI Director Jim Comey**
4. **The Logjam (and Another) Vulnerability against Diffie-Hellman Key Exchange**

Filed under: Cyber & Technology, Cybersecurity, Unfiled

CRAFTED BY CHARLIE ON WORDPRESS ¶ INSPIRED BY VERYPLAINTXT BY SCOTT WALLICK ¶ © 2015