acmqueue

Resolved: the Internet Is No Place for Critical Infrastructure

Risk is a necessary consequence of dependence

Dan Geer

What is critical? To what degree is critical defined as a matter of principle, and to what degree is it defined operationally? I am distinguishing what we say from what we do.

Mainstream media love to turn a spotlight on anything they can label "hypocrisy," the Merriam-Webster unabridged dictionary meaning of which is:

the act or practice of pretending to be what one is not or to have principles or beliefs that one does not have, especially the false assumption of an appearance of virtue

The debate topic I am proposing can therefore be restated as calling out, "Hypocrisy!" on the claim that the Internet is a critical infrastructure either directly or by transitive closure with the applications that run on or over it. If the claim were true, the divergence between our beliefs and our practices would be necessarily narrower (by *our* I mean each of us both separately and collectively).

Perhaps I am echoing how a free-range cattleman felt about the coming of barbed wire, roads, and land title to the American West. The great cattle drives of the West lasted 20 years before other kinds of progress made them impossible. Commercial Internet traffic began some 20 years ago, with the interconnection of PSInet and UUNet by CIX (Commercial Internet Exchange).

Recalling Winston Churchill's "The further back I look, the further forward I can see," either the wide open range that is the freedom of an Internet built on the end-to-end principle must die, or else we must choose not to allow the critical infrastructure of our lives to depend on that Internet. Freedom and reliability are now at odds.

Consider the Internet as a Hobson's choice: either you get it, warts and all, or you get nothing. According to The Pew Research Center's Internet and American Life Project:[13]

One in five American adults do not use the Internet. Among adults who do not use the Internet, almost half [said] that the main reason they don't go online is because they don't think the Internet is relevant to them.

For those 10 percent who, presented with a take-it-or-leave-it proposition regarding the Internet, choose "leave it," the Internet does not register as desirable and may, for some of them, be undesirable.

OPTING OUT IS HARDLY AN OPTION

I have never bought or owned a television. There is no social opprobrium if you opt out of television; it is merely a choice. That 10 percent of the population that doesn't bother with the Internet is surely similar to whatever fraction of the population doesn't see any reason to bother with television. But can they refuse the Internet and have that just be something they choose not to do anything with, and therefore be inconsequential to their lives, the way television is inconsequential to mine?

No. It is not possible to live your life without having a critical dependence on the Internet, even if you live at the end of a dirt road but still occasionally buy nails or gasoline. Unlike television, you cannot unplug from the Internet even if you want to. If you are dependent on those who are dependent on television, then so what? If, however, you are dependent on those who are dependent on the Internet, then so are you. Dependence with respect to television is not transitive. Dependence with respect to the Internet is.

Because dependence on the Internet is transitive, those who choose to "leave it" are still dependent on it unless they are living a pre-industrial life. That rejectionists depend on people who are not rejectionist is simply a fact.

But rejectionists do have impact—they are now a kind of fail-safe. If we begin to penalize the rejectionists—that is, force them to give up on their rejectionism—we will give up a residuum of societal resiliency.

To illustrate, I have a 401(Kk) account with Fidelity Investments. Fidelity no longer accepts client instructions in writing; it only accepts instructions over the Internet or, as a fallback for the rejectionist, over the phone. It simply does not accept the canonical wet ink signature on bond paper. I have sent Fidelity paper letters, and its representatives have responded in e-mail messages that say just that (though I should note that I never gave them my e-mail address). Fidelity's stand is that its auditors approve of this scheme. My stand is, "Your auditors work for you, not me." Those e-mail letters do not contain a digital signature and, in any case, what is the equivalent of that for a phone call? Fidelity still sends paper statements to the same mailing address from which I have been writing.

I use a small local bank. I sent it a letter stating that as I would not be using online services, I would like the bank to turn off access to my account and raise an alarm if anyone ever tried to use the uninitialized account waiting in my name. The bank agreed without any argument. That is not the norm. Try, as I have done, to make that same request to the arm of the payroll services giant ADP that runs the get-your-W2-online service called iPay. It will refuse.

Estonia is the most Internet-dependent country,[7] and Estonian pride is entirely in order. Its degree of dependence happens not to be for me: I want to retain the ability to opt out of direct dependence on the Internet—that is, to opt out of that dependence that is the root of risk. I mean that as stronger than a preference but weaker than an ultimatum.

In a free society, that which is not forbidden is permitted. In a non-free society, that which is not permitted is forbidden. Obamacare deploys the government's monopoly on the use of force to collectivize the downside risk of illness. Just as forcibly collectivizing the downside risk of illness has its utopian proponents, so, too, does forcibly collectivizing the downside risk of Internet exposure.

Estonia is well ahead of nearly everybody in intentional dependence on the Internet; China is well ahead of nearly everybody in forcibly collectivizing the extent and manner in which the Internet is used by Chinese citizens. As sovereigns, the former is Estonia's right just as the latter is China's right. I want neither, even though I must acknowledge that as nations decide on their particular mix of dependencies, the Internet will be dramatically balkanized. The Internet will never again be as free as it is today.

In 2002 a total computer outage occurred at Harvard's Beth Israel Hospital.[6] The event was severe and unexpected, and recovery was frustrated by complexity. That a fallback to manual systems was possible saved the day, and it was those who could comfortably work without network dependence

who delivered on that possibility, because they were old enough to have done so at earlier times.

Risk is a consequence of dependence. Because of shared dependence, aggregate societal dependence on the Internet is not estimable. If dependencies are not estimable, then they will be underestimated. If they are underestimated, then they will not be made secure over the long run, only over the short. As the risks become increasingly unlikely to appear, the interval between events will grow longer. As the latency between events grows, the assumption that safety has been achieved will also grow, thus fueling increased dependence in what is now a positive feedback loop. Accommodating rejectionists preserves alternative, less complex, more durable means and therefore bounds dependence. Bounding dependence is the core of rational risk management.

COMMON-MODE FAILURE

In the language of statistics, *common-mode failure* comes from underappreciated mutual dependence. Quoting the National Institute of Standards and Technology:[9]

[R]edundancy is the provision of functional capabilities that would be unnecessary in a fault-free environment. Redundancy is necessary, but not sufficient for fault tolerance.... System failures occur when faults propagate to the outer boundary of the system. The goal of fault tolerance is to intercept the propagation of faults so that failure does not occur, usually by substituting redundant functions for functions affected by a particular fault. Occasionally, a fault may affect enough redundant functions that it is not possible to reliably select a non-faulty result, and the system will sustain a common-mode failure. A common-mode failure results from a single fault (or fault set). Computer systems are vulnerable to common-mode resource failures if they rely on a single source of power, cooling, or I/O. A more insidious source of common-mode failures is a design fault that causes redundant copies of the same software process to fail under identical conditions.

That last part—"A more insidious source of common-mode failures is a design fault that causes redundant copies of the same software process to fail under identical conditions"—is exactly the kind of fault which can be masked by complexity, precisely because complexity ensures underappreciated mutual dependence.

Which brings us to critical infrastructure and the interconnection between critical infrastructures by way of the Internet. Quoting the Clinton Administration's Policy on Critical Infrastructure Protection from May 22, 1998:

Critical infrastructures are those physical and cyber-based systems essential to the minimum operations of the economy and government.[11]

"Essential to minimum operations" does not imply a requirement that the armor deflect all bullets, only that no bullet is paralyzing. One of the great Allied victories of World War II was getting 338,000 soldiers off the beaches of Dunkirk using 800 "little boats," a paragon of the phrase "essential to minimum operations."

The Internet is a network of networks, its main protocols designed for tolerance to random faults and for the absence of common-mode failure. It has been proven in practice.[2] It was not designed, however, for resistance to targeted faults, which cannot be done at the same time as you are designing for resistance to random faults.[1]

In an Internet crowded with important parts of daily life, the chance of common-mode failure is no idle worry. The Obama administration is broadly committed to increasing dependence on the Internet, most notably on two fronts: electronic health records and the so-called Smart Grid, either of which might be said to be "essential to the minimum operations of the economy and government." As with most garden paths, both can have eminently useful results for which a desire is rational. Both illustrate my point.

Electronic health records depend on the smooth functioning of electric power, networks, computers, displays, and a host of security features particularly as they relate to maintaining consistency across multiple practices.[12] The Smart Grid depends on almost everything we now know about power, including the absolute necessity of good clocks, a wide range of industrial controls operated flawlessly at distance and guaranteed not to lie about their state, and another host of security features. Both of these involve new levels of exposure to common-mode risk. Doing without their benefits will be easier for those who can remember not having had them.

Each new dependence raises the magnitude of downside risk, the potential for collateral damage, and the exposure of interrelationships never before realized. Forget the banks, it is the Internet that is too big to fail. While there is no entity that can bail out the Internet, there is no developed country that is not developing ways to disrupt the Internet use of its potential adversaries. The most a country can count on being able to do is preserve the Internet interior to itself—as Estonia demonstrated when under attack from Russia—though at some level of transborder interconnection, the very concept of "interior" loses meaning.

Designing for tolerable failure modes is precisely what security engineering is fundamentally about. The failure mode you did not think of will not be in your design; therefore, whether it is tolerable will depend on other things. The question, then, is, Can tolerable failure modes be designed—that is to say, Can a failure mode never before possible be added to the system such that larger, intolerable failures can be precluded? Is there a cyber-critical infrastructure analog to a shear-bolt in the drivetrain of heavy machinery?

No country, no government, no people need rules against things that are impossible. Our onrushing dependence on things never before possible creates vacua where, in the fullness of time, there will have to be rules. As Chicago Mayor Rahm Emanuel put it, the creation of rules is easier in a time of crisis, so one must "never let a good crisis go to waste." He is right as a matter of observation; he is wrong as a matter of probity. Just as driving under the influence of alcohol is wrong, so is making policy under the influence of adrenaline. Eleven years before he became the fourth president of the United States, James Madison said:

"Perhaps it is a universal truth that the loss of liberty at home is to be charged to provisions against danger, real or pretended, from abroad."

One wonders how Madison would feel about an interconnected world where *abroad* has so thoroughly lost its meaning, at least with respect to Internet-dependent critical infrastructure if not national frontiers. My guess is that Madison would decide that the Internet is, per se, "abroad." As such, our critical infrastructure is now another country, something from which to be protected at the cost of loss of liberty.

I've previously spoken on whether having people in the loop for security is a fail-safe or a

liability.[3] I won't recount the arguments here, but I will give my conclusion: a good security design takes people out of the loop except when it cannot and, when it cannot, makes clear that this is so. Putting a human in the loop—that is to say, falling back from automation—has proven to be a breakthrough finesse.

That the public has "volunteered" its unused computing power to botmasters is a historical mirror of how press gangs once filled the rosters of the British Navy, but how is that meaningfully different from a formal mandate that if you have medical records those shall be electronic, or if you receive electricity that the meter be a surveillance tool? How is it different from finding that compliance auditors have certified to distant regulators that there is no need to accept a signed paper letter detailing the wishes of the financial client?

WHAT PRICE SECURITY?

Security is a necessary but insufficient condition for reliability. As such, connecting the insecure (and thus unreliable) to the important and expecting the mélange to be reliable is utter foolishness. As network security expert Marcus Ranum says, "A system that can be caused to do undesigned things by outsiders is not 'reliable' in any sense of the word." Work being done by Sergey Bratus, Meredith Patterson, and others at LANGSEC (Language-theoretic Security) yields insight deserving full quotation:[8]

The Language-theoretic approach regards the Internet insecurity epidemic as a consequence of ad hoc programming of input handling at all layers of network stacks, and in other kinds of software stacks. LANGSEC posits that the only path to trustworthy software that takes untrusted inputs is treating all valid or expected inputs as a formal language, and the respective input-handling routines as a recognizer for that language. The recognition must be feasible, and the recognizer must match the language in required computation power.

When input handling is done in [an] ad hoc way, the de facto recognizer, i.e., the input recognition and validation code ends up scattered throughout the program, does not match the programmers' assumptions about safety and validity of data, and thus provides ample opportunities for exploitation. Moreover, for complex input languages the problem of full recognition of valid or expected inputs may be UNDECIDABLE, in which case no amount of input-checking code or testing will suffice to secure the program. Many popular protocols and formats fell into this trap, the empirical fact with which security practitioners are all too familiar.

Viewed from the venerable perspective of Least Privilege, ... computational power is privilege, and should be given as sparingly as any other kind of privilege to reduce the attack surface. We call this ... the Minimal Computational Power Principle.

We note that recent developments in common protocols run contrary to these principles. In our opinion, this heralds a bumpy road ahead. In particular, HTML5 is Turing-complete, whereas HTML4 was not.

Almost nothing we have in our cyber interfaces to critical infrastructure meets LANGSEC's test. Attaching the cyber interface of critical infrastructure to the Internet is a flat-out guarantee of error. Such error may be improbable, but probabilistic events eventually occur. If we are especially unlucky, those errors will not be prompt.

There has been much talk about whether to grant the President a kill switch for the Internet.

There is some logic to that if, due to interdependence that is inestimable, it is not possible to disambiguate friend from foe. Were someone on an inbound airplane found to have smallpox, the passengers and crew would be quarantined as a matter of public health until each of them could be separately certified as disease-free. Many important enterprises, public and private, quarantine inbound e-mail with nearly as much vigor as they quarantine inbound DHL packages. The logic is sound. The time scale is human.

We have amongst ourselves, and we accommodate, cloistered communities such as the Amish. If a food crisis were to materialize, it is the Amish who would be least affected. We also have amongst ourselves neo-Luddites, who know where machines will lead and on that basis may well mimic their progenitors. The Amish merely wish to be left alone. Is there not room in our increasingly wired world for those who choose merely to be left alone, in this case choose not to participate in the Internet society? Do those who do not participate deserve to not have their transactions of all sorts be exposed to a critical infrastructure dependent on the reliability of Internet applications?

The United States' ability to project power depends on information technology, and, as such, cyber insecurity is the paramount national security risk.[4] Putting aside an Internet kill switch, might it be wise for the national authorities to forbid, say, Internet service providers from propagating telnet, SSH v1, or other protocols known to be insecurable? If not, should cyber components of the Defense Industrial Base be forbidden to accept such connections? There is a freedom-vs.-reliability collision in that—if not a natural policy. There is a direct historical echo as well; in 1932 the foremost political commentator of the age, Walter Lippmann, told President Roosevelt, "The situation is critical, Franklin. You have no alternative but to assume dictatorial powers."

Again, when 10 percent of the population sees nothing in the Internet for them, should we respect that wish and ensure that, as with the Amish, there is a way for them to opt out without having to live in a cave? Should we preserve manual means for them?

I say yes, and I say so because the preservation of manual means is a guarantee of a fallback that does not have a common-mode failure with the rest of the interconnected, mutually vulnerable Internet world.

My colleague and I run the Index of Cyber Security.[5] Our respondents all have direct operational responsibility for cyber security. The Index is rising—that is, experts say risk is accumulating in much the same way that burnable timber accumulates on the eastern slope of the Rockies. This is a formal, metrics-based backstop to saying that "we" are not running fast enough to stay in the same place; therefore, preserving fallback is essential.

Department of Defense thinkers agree; their goal is no longer intrusion prevention but intrusion tolerance. If we are to practice evidence-based medicine on the body Internet, we must first acknowledge that expensive therapy is not always the answer. Cost-effective medicine cannot be practiced if every human life is infinitely valuable. Perhaps you can come up with a cyber analog to "quality-adjusted life years" and help us all decide when to treat, when to palliate, and when to accept mortality.

The following ideas from the Homeland Security Watch blog may be ones whose time has come:[10]

The deficient and unforgiving design that many of us—private citizens, as well as public safety agencies—have adopted is dependence on just-in-time information.

My twentysomething children seldom preplan in any significant way. They expect cell phones, text messaging, Facebook, and e-mail to allow them to seize the best opportunities that unfold. It works and I envy them. Except when it does not work. Except when these digital networks fail.

Much of our consumer culture is built around the same approach. We have become an economy, a society optimized for just-in-time. It can be a beautiful dance of wonderful possibilities emerging in a moment and rapidly synchronized across time and space. Until the music stops.

...There is a shared overconfidence in the fail-safe capabilities of protective design and effective communications. [T]he design bias increase[s] risk exposure, communications was confusing or worse, and both the design and the communications protocols complicate effective human response once risk [is] experienced.

SUMMING UP

Risk is a consequence of dependence. Because of shared dependence, aggregate societal dependence on the Internet is not estimable. If dependencies are not estimable, then they will be underestimated. If they are underestimated, then they will not be made secure over the long run, only over the short. As the risks become increasingly unlikely to appear, the interval between events will grow longer. As the latency between events grows, the assumption that safety has been achieved will also grow, fueling increased dependence in what is now a positive feedback loop. If the critical infrastructures are those physical and cyber-based systems essential to the minimum operations of the economy and government, and if leading cyber-security operational management says risk is growing steadily, then do we divert more of our collective power to forcing security improvements that will be sharply diseconomic, or do we preserve fallbacks of various sorts in anticipation of events that seem more likely to happen as time passes?

Does "use it up, wear it out, make it do, or do without" have any meaning for us? Is centralizing authority the answer, or is avoiding further dependence the better strategy? Can we imagine starting over in any real sense, or is balkanization not just for nations but for critical sectors as well? Is the creative destruction that is free enterprise now to be focused on remaking what are normally the steadying flywheels of American society, by which I mean government and other capital-intensive industries? Do we celebrate the individual who still prefers to fix things he or she already has, or are those individuals to be herded into national health information networks, Smart Grids, and cars that drive themselves?

REFERENCES

1. Barabasi, L., Albert, R. 1999. Emergence of scaling in random networks. *Science* 286 (October): 509-512.
2. Branigan, S., Cheswick, B. 1999. The effects of war on the Yugoslavian Network. http://www.cheswick.com/ches/map/yu/index.html.
3. Geer, D. 2012. People in the loop: are they a fail-safe or a liability? Suits & Spooks (February 8); http://geer.tinho.net/geer.suitsandspooks.8ii12.txt.
4. Hathaway, M. 2009. Securing our digital future. White House blog.; http://www.whitehouse.gov/blog/2009/05/29/securing-our-digital-future.

5. Index of Cyber Security; http://cybersecurityindex.org.
6. Kilbridge, P. 2003. Computer crash—lessons from a system failure. *New England Journal of Medicine* 348(10): 881-882; http://ehealthecon.hsinetwork.com/NEJM_downtime_2003-03-06.pdf.
7. Kingsley, P. 2012. How tiny Estonia stepped out of USSR's shadow to become an Internet titan. *The Guardian*; http://www.guardian.co.uk/technology/2012/apr/15/estonia-ussr-shadow-internet-titan.
8. LANGSEC: Language-theoretic Security, http://langsec.org.
9. National Institute of Standards and Technology, High Integrity Software System Assurance. 1995. Redundancy management (section 4.2). A Conceptual Framework for System Fault Tolerance; http://hissa.nist.gov/chissa/SEI_Framework/framework_16.html.
10. Palin, P. 2012. Can you envision a "successful failure"? Homeland Security Watch; http://www.hlswatch.com/2012/07/13/can-you-envision-a-successful-failure/.
11. Presidential Decision Directive 63. 1998. The Clinton Administration's Policy on Critical Infrastructure Protection.; http://www.fas.org/irp/offdocs/paper598.htm.
12. Shim, S. S. Y. 2012. The CAP theorem's growing impact. *IEEE Computer* 45 (2): 21-22.
13. Zickuhr, K., Smith, A. 2012. Digital differences. Pew Research Center; http://pewinternet.org/~/media//Files/Reports/2012/PIP_Digital_differences_041312.pdf.

LOVE IT, HATE IT? LET US KNOW

feedback@queue.acm.org

DAN GEER is a security researcher with a quantitative bent. His group at MIT produced Kerberos, and after time at a number of startups, he is still at it—today as CISO at In-Q-Tel. He writes a lot at every length, and sometimes it gets read, such as the semi-famous paper on whether a computing monoculture rises to the level of a national security risk. He's an electrical engineer, a statistician, a farmer, and someone who thinks truth is best achieved by adversarial procedures.