Metrics, Economics, and Shared Risk at the National Scale

Dan Geer
dan@geer.org / 617.492.6814

formalities: Daniel E. Geer, Jr., Sc.D.

Principal, Geer Risk Services, LLC
P.O. Box 390244
Cambridge, Mass. 02139
Telephone: +1 617 492 6814
Facsimile: +1 617 491 6464
Email: dan@geer.org

VP/Chief Scientist, Verdasys, Inc.
950 Winter St., Suite 2600
Waltham, Mass. 02451
Direct-in: +1 781 902 5629
Corporate: +1 781 788 8180
Facsimile: +1 781 788 8188
Email: geer@verdasys.com

Outline

- Where are we?
- What drives change?
- The nature of risk
- The near term future
- Measurement, models, implications
- Summary and proposal

An general thread of the thoughts in this presentation.

Ask the right questions

(What can be more engineering-relevant than getting the problem statement right?)

- What can attack a national infrastructure?
- What can be done about it?
- How much time do we have?
- Who cares and how much do they care?

In all of engineering, getting the problem statement right is job 1. Without the right problem statement you get "we solved the wrong problem" or "this is a solution is search of a problem" or worse.

Our questions here are to ask what it is about the national scale that elevates some attacks to proper focus and what sets others aside.

The Setting

- Advanced societies are more interdependent
- Every sociopath is your next door neighbor
- Average clue is dropping
- Information assets increasingly in motion
- No one owns the risk -- yet

The more advanced the society the more interdependent it is. Which is the cause and which is the effect is a debate for sociology or economics, but it is a tight correlation.

Equidistance and near zero latency is what distinguishes the Internet from the physical world.

Power doubles every 12-18 months and, obviously, skill on the part of the user base does not. Hence the ratio of skill to power falls. This has broad implications.

Information does not want to be free, but it does want to be located to its advantage.

In finance, risk taking and reward are tightly correlated and there is zero ambiguity over who owns what risk; cf., in the digital security sphere where there is nothing but ambiguity over who owns what risk.

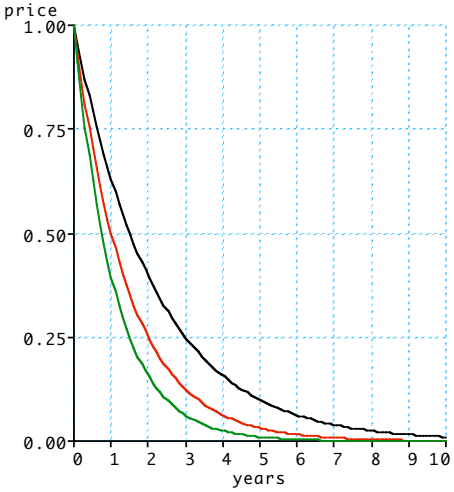
The Drivers of Change

- Laboratory
- Economics
- Psychology

What is it that changes natures of the computing infrastructure at the national level? For relevance to decision making at that level, it essential to look not at the present moment but rather what trends exist extrapolated to at least that point in the future which is the earliest practical time at which strategic countermeasures can intercept the threat to the national infrastructure. Put differently, as one cannot expect to turn a ship the size of the national infrastructure in short time we must therefore lead our target.

There are three principal drivers to the national computing infrastructure: the ongoing miracles exiting our commercial laboratories, the economics by which change in our national infrastructure are modified, and the psychology of national populations, generally speaking, which latter point determines what it is that the public demands of government, inter alia.

Lab: model creep



ten years
CPU 10^2
Data 10^3
BW 10^4

Black line is “Moore’s Law” whereby $/MHz drops by half every 18 months. It’s unnamed twins are, in red, the price of storage (12 month) and, in green, bandwidth (9 month). Taken over a decade, while CPU will rise by two orders of magnitude, the constant dollar buyer will have 10 times as much data per computer cycle available but that data will be movable to another CPU in only 1/10th the time. This has profound implications for what is the general charactgeristic of the then optimal computing plant.

And, even if there are wiggles here and there, the general point that there is a drift over time in the optimal computer design stands.

Econ: applications

- Applications are federating, and thus
  - accumulating multiple security domains
  - getting ever more moving parts
  - crossing jurisdictions

Under economic influences, such as the various promises of "web services," applications in general are increasing their reach by federating across internal and external corporate boundaries not to mention jurisdictions. That this requires more moving parts is obvious, of course.

This force is not the issue, its effect is. The effect is to make ever-larger applications at least insofar as these ever-larger applications are able to be productivity-enhancing while exhibiting complexity-hiding.

Econ: transport

- HTTP assumes transport role, and thus
  - attack execution at lower skill levels
  - content inspection collapses
  - perimeter defense trends diseconomic

Allied with the increasing reach and scope of applications is an increasing reliance on HTTP as the transport mechanism. Microsoft, for its .NET environment, is actually recommending that application writers focus on libhttp rather than libtcp, i.e., to rely on HTTP as the core transport infrastructure rather than TCP.

As the limit, a firewall needs one hole and only one hole — for HTTP (and HTTPS, i.e., SSL). With a hole in the firewall of this size, it is hardly worth saying the level of effort and skill required to transit the firewall to attack internal machines is lessened. Which is more, once program fragments are part of the payload (such as remote procedure calls in the Simple Object Application Protocol (SOAP)), content inspection of the information flow becomes virtually intractable.

Econ: data

- Data takes command, because
  - corporate IT spending on storage:
    4% in 1999 v. 17% in 2003 (Forrester)
  - data/$ up 16x in same interval
  - total volume doubling at ~30 months

The volume of data is substantial, getting more so, and will likely dominate security's rational focus from this point forward.

The public's interest

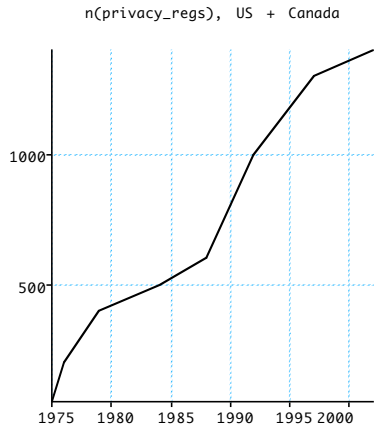
- Spam
  - channel saturation, labor costs
- Viruses
  - warning-time shrinking, labor costs
- Theft
  - identity, cycles, keystrokes, reputation

...Safety, safety, safety

The interest of individual members of the public includes these illustrative three, at least.

One reaction

1975	50
1976	200
1979	400
1984	500
1988	600
1992	1000
1997	1300
2002	1400



One measure of the Public Interest is the rate at which the security setting, in this case its subset privacy, is regulated.

This graph and data are of the total number of privacy regulations at the State and Federal level in the US plus Canada.

The Public Interest

- Loss of inherently unique assets
  - GPS array, FAA EBS, DNS
- Cascade failure
  - Victims become attackers at high rate

Everything else is less important

If having to name the only risks that matter at the national scale, there seem to be two classes and only two classes.

On the one hand, there are entities that are inherently unique by design. For example, the Global Positioning System satellite array (taken as a unit) is one such entity; the Federal Aviation Administrations emergency broadcast system is another, and the Domain Naming System is another. In each case, it is an authoritative data or control source which would be less authoritative if it was surrounded by alternatives. Putting it differently, you only want one red telephone though losing that red telephone is a risk at the national scale.

On the other hand, there entities that are dangers to each other in proportion to their number -- any risk which, like an avalanche, can be initiated by the one but propagated by the many. This “force multiplication” makes any of this class of risks a suitable candidate for national scale.

Risk to Unique Asset

- Pre-condition: Concentrated data/comms
- Ignition: Targeted attack of high power
- Counter: Defense in depth, Replication
- Requires: The resolve to spend money

For unique assets to be a risk at the national scale, you need the pre-condition of some high concentration of data, communications, or both. The ignition of that risk is a targeted attack of high power up to and including the actions of nation states. The counter to this latent risk is "defense in depth" which may include replication. Defense in depth is ultimately (at the policy level) a referendum on the willingness to spend money.

As such, there is nothing more to say at the general level and we lay this branch of the tree aside so as to focus on the other.

Risk of Cascade Failure

- Pre-condition: Always-on monoculture
- Ignition: Any exploitable vulnerability
- Counter: Risk diversification, not replication
- Requires: Resolve to create heterogeneity

For cascade failure to be a risk at the national scale, you need the pre-condition of an always-on monoculture. The ignition of that risk is an attack on vulnerable entity within the always on monoculture so long as it has a communication path to other like entities. The counter to this latent risk is risk diversification which absolutely does not include replication. Cascade avoidance is ultimately (at the policy level) a referendum on the resolve to treat shared risk as a real cost, per se.

We now follow this branch to see where it leads. Sean Gorman of George Mason University has an upcoming publication that suggests that the risk-cost of homogeneity kicks in at rather low densities (preliminary results indicate 43% for leaf nodes, 17% for core fabric).

Why 'sploits matter

- Monoculture is a force multiplier
- Amateurs provide smokescreen for pros
- Only known vulns get fixed
  - The unknown are held in reserve
- Automated reverse engineering of patches is accelerating

So why do exploits matter? Because in a monoculture they are the ignition and their propagation amongst potential instigators of a cascade failure is well documented. Of course, the extent of their existence and propagation is unknowable in and of itself, but it is clear that the testing of exploits by the most expert is sufficiently obscured by the constant rain of amateur attacks. One estimate (by John Quarterman of Internet Perils) is that perhaps 10% of total Internet backbone traffic is low-level scans while another (by Vern Paxson of Lawrence Berkeley) is that for a site such as LBL one can expect perhaps 40% of inbound connections to be attacks.

Because only known vulnerabilities get fixed, the central question is who knows what and when. The conservative assumption for a vulnerability discoverer is that he was not the first to discover the current vulnerability. A similarly conservative assumption is that not all vulnerability discoverers are of good will. Therefore the question is "How many vulnerabilities are known, silently, to persons not of good will?" The corroborating evidence that this number is non-zero lies in observing that all major virus or worm attacks to date have exploited previously known vulnerabilities, never unknown ones. With such evidence, either all vulnerabilities are discovered by persons of good will or there is a reservoir of vulnerabilities being held in reserve.

Note that converting patches into vulnerabilities by reverse engineering the patches is not only now the dominant source of exploits but that it is becoming much quicker due to automation. In two years the times have dropped from six months to under a week for principal attacks of public interest. Further declines may no longer matter.

Wishful Thinking

- The absence of a serious event can be:
  - Evidence of zero threat
  - Consistent with risk aggregation
  - A failure to detect precursors

The absence of a major attack event, the situation in which we find ourselves today, is not, as it might seem on first blush, reassuring of low threat. It is consistent with low/no threat to be sure, but it is also consistent with risk aggregation (an insurance term where instead of 1,000 claims occurring at random one instead gets 1,000 claims all at once, the difference between house fires and earthquakes). It is also consistent with a failure to detect, though less likely that “major” and “indetectible” are likely to appear in the same sentence.

Microsoft in particular

- Tight integration as competitive strategy
  - users locked-in
  - effective module size grows
  - reach of vuln expands
- Insecurity α complexity

The situation with Microsoft is the critical focus just as when discussing solar power one must speak of Sol.

The quality control literature leads one to expect that as effective code size grows complexity grows as the square of that code size. Similarly, the quality control literature expects total flaws, of which security flaws are a subset, to grow linearly in complexity. Microsoft's competitive strategy is manifestly to achieve user-level lock-in via tight integration of applications. This tight integration, besides violating software engineering wisdom, expands effective module size to that of the tightly integrated whole and thus inevitably creates the platform most likely to have security flaws, and by a wide margin. Coupled with its 94% market share one thus achieves the vulnerable monoculture on which cascade failure depends.

Software

"[B]y using this product you agree that it's all your fault, that it's only broken to the extent that it ships 'as is' and therefore if you think it's broken you accepted that this was the case when you bought it, and anyway you agreed it wasn't and you didn't buy it anyway, because it's still ours..."

http://www.theregister.co.uk/content/4/33082.html

This is the wonderfully curmudgeonly UK digital publication "The Register" synopsizing the plain english meaning of most software licenses.

If nothing else, it illustrates the lack of clarity about responsibility when software is misused.

Field Repairs

- Not possible to patch your way to safety
- Liability will reference patch-state
  - "Due care" vs "Force Majeure"
  - "Attractive nuisance" vs "Unwitting accomplice"
- Automatic update is the most powerful form of mobile code

Field repairs, the dominant activity if not strategy of the present time, are at best a damage containment. It is not possible to patch yourself to safety: any significant patch latency preserves the critical mass (of vulnerable entities) while every patch has a non-zero chance of collateral damage. Thus we come to a discussion of formal liability and the high likelihood that in the short term such discussion will focus on patch-state as a proxy for culpability either in the sense of patching being evidence of due care or the lack of patching, particularly within substantial enterprises, being evidence of an attractive nuisance (like an unfenced swimming pool).

Looking dispassionately at risk, one must also conclude that automatic update is the ultimately powerful form of mobile code. Automatic patching does harden systems but it does so more toward brittleness than toward toughness in that if ever the automatic patching pathway is itself effectively co-opted then the game is largely over at that moment.

Prediction(s)

- Traffic analysis recapitulates cryptography
- Perimeter defense contracts to data
- Security & Privacy have their long-overdue head-on collision
- Meritocracy begins yielding to government

No discussion of national level threat can look at the current point in time; it must instead lead its target just as a hunter must his. In that sense, the next ten years (or less) will have the commercial sector catching up to the military in traffic analysis just as the last ten years had that catch-up in cryptography. At the same time, increasing threat will, as it must, lead to shrinking perimeters thus away from a focus on enterprise-scale perimeters and more toward perimeters at the level of individual data objects. Security and privacy are, indeed, interlocking but, much as with twins in the womb, the neoplastic growth of the one will be to the detriment of the other hence the bland happy talk of there being no conflict between the two will be soon shown to be merely that. Finally, the Internet as a creature built by, of, and for the technical and ethical elite being no longer consistent with the facts on the ground, its meritocratic governance will yield to the anti-meritocratic tendencies of government(s).

Grand Challenges

...within ten years...

- No further large scale epidemics
- COTS tools for building certifiable systems
- Low/no skill required to be safe
- Info. risk mgmt. ≥ financial risk mgmt.

In November, 2003, the Computing Research Association held a limited attendance, invitation only retreat in Virginia at the behest of the National Science Foundation. The purpose was to set the ten-year research agenda in information security <http://www.cra.org/Activities/grand.challenges/security/home.html>. Here are the results in lay terms: An end to epidemics, commercial off the shelf (COTS) tools for building certifiable systems, improvements in semantics and user interface such that one need not be an expert to be safe, and information risk management of a quantitative sophistication as good as that of financial risk management.

These are high goals, and at the same time it is horrifying that any of them could take a decade to deliver. On the other hand, if they do take as much as a decade, then starting now is crucial.

See http://www.cra.org/Activities/grand.challenges/security/home.html

Metrics

"When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science."

William Thomson, Lord Kelvin

The foremost statement in science regarding the necessity of measurement.

Metrics: our version

- How secure am I?
- Am I better off than this time last year?
- Am I spending the right amount of money?
- How do I compare to my peers?
- What risk transfer options do I have?

These are precisely the questions that any CFO would want to know and we are not in a good position to answer.

Metrics: lay of the land

- No time to create hence **adapt**
- Information sharing, various forms
- To be relevant, must calibrate spend
  - GAAP for security a long way off
- et cetera, et cetera, et cetera

The need for security is so great that we simply cannot afford to wait while new measures, uniquely wonderful and wonderfully unique, are invented. We must adapt measures and techniques that already exist. Yes, this requires taste and judgment, but those, however rare, are in better supply than time.

Principal amongst the measurement efforts must be some way to draw baselines and otherwise to share data such that individual firms can compare themselves to others. The information to share is not exception data but ordinary data. That which is un-ordinary cannot be described much less identified until that which is ordinary is described first. To take a trivial example, without sharing firewall logs with like firms you cannot know whether you are the recipient of attacks that are just like everyone else's (thus making you a target of chance) or that you are a recipient of purpose-built attacks aimed just at you (thus making you a target of choice).

Whatever we measure, it has to be valuable as a mechanism to calibrate spending. Unfortunately for industry, a set of generally accepted accounting principles may be in place for finance but it is far from in place for security.

Metrics: adapt not create

- Beg, borrow, and steal from
  - Public health (CDC)
  - Accelerated failure time testing (MTTR)
  - Insurance (ALE, Cat Bonds)
  - Portfolio management (VAR)
  - Physics (scale-free networks)

The future belongs to the quants, full stop. As such, and bearing in mind the critical need for security at this time, we must borrow from other fields as we have no time to invent everything from scratch. We are likewise lucky that at this time the field has the maximum of hybrid vigor in that all of its leaders were trained at something else hence our ability to extract from that "something else" is maximal.

For a sense of this, see <http://www.stake.com/research/reports/acrobat/ieee_quant.pdf>, or <http://www.securitymetrics.org>.

Metrics: info sharing

- Models for information sharing
  - Central, mandatory, top-down
  - Specific, enlightened self-interest
  - Exception vs routine

In a shared infrastructure at the national scale, some metrics will necessarily concern themselves with shared data. The nature of shared data is the focus of the next three slides, viz., whether that data is mandatorily with a central authority, shared amongst like entities on the basis of enlightened self interest, or whether some other form of sharing is appropriate and, if so, whether the data shared is routine data or exception data.

Sharing: top down

- Centers for Disease Control
  - Mandatory reporting of communicable diseases
  - Longitudinal analysis of incidence and prevalence (with lab confirmation)
  - Away teams to handle outbreaks (hemorrhagic fevers like Ebola)

The United States Centers for Disease Control play a global role, and no public health practitioner fails to read the weekly Mortality and Morbidity Report. In a paper published in the Proceedings of the (August 2002) USENIX Security Symposium, Staniford, et al., proposed a CDC for the Internet. It would parallel the role of the existing CDC in that it would enjoy mandatory reporting of communicable diseases, it would perform longitudinal analysis of incidence and prevalence of disease including the search for excess incidence or prevalence, and it would have away-teams to handle outbreaks of disease.

The CDC are established by the rule of law and paid for by the rule of law (through taxes). They were not present at the creation, but are now essential. When, if at all, should the rule of law include mandatory reporting, forced treatment and/or quarantine of the ill, the publication of an Internet-equivalent of the MMR, formal predictive work that aids those planners who must anticipate epidemics whether of flu or NIMDA, and the public identification of locales, however defined, where there is an excess of incidence or prevalence of any particular pathogen.

Sharing: semi-self-interest

- Information Sharing & Analysis Centers
  - PDD 63 (1998), sector-specific
  - Report to relevant dept (FS to Treasury)
  - Explicit exemption from FOIA, anti-trust
  - Anonymous submission, sort of
  - DHS wants them all

The Information Sharing and Analysis Centers were created under Presidential Decision Directive #63, Clinton, 1998. Each ISAC is sector specific and reports to a sector coordinator. "Sector specific" means financial services, information technology, energy, electric power, and so forth. All are voluntary associations of private and public firms acting on behalf of their individual sectors. The sector coordinator will report in to a relevant cabinet-level department, e.g., Treasury for the Financial Services ISAC. Data sharing and other interactions with the ISACs enjoy explicit exemption from the Freedom of Information Act and from Anti-Trust. They share some data in a titularly anonymous way that is more like unattributed than anonymous (there being only guarantees of anonymity rather than technical means). The Department of Homeland Security, Infrastructure Analysis and Protection Authority, wants to take them all over. At issue is whether the private sector of the US economy, which owns 90+% of the nation's critical infrastructure can, will, or will be allowed to ensure its protection from shared risks. This is as close to a "command economy" move as one is likely to see from any administration for some time. The ISACs need to lead or follow.

Disclaimer: The present author serves as one of five outside advisors to the FS/ISAC, pro bono.

Sharing: regulation

- Information sharing of a sort
  - Disclosure (Calif SB1386, FTC, ...)
  - Outlier-focus distorts understanding
  - Government cannot be expected to hold anything confidential

Regulatory sharing is sharing of a sort. Looking forward, it is clear that security-related regulation, broadly defined, is here to stay. The most likely area of regulation to develop first and most fully is that around disclosure.

With disclosure, there is already the example of California's Senate Bill #1386 (SB1386) which mandates disclosure when a firm may have lost personally identifiable data entrusted to it by its clients. (This has the quality of "proving a negative" in that the disclosure is required when it cannot be proved that the data is still safe.) Similar laws have since been adopted, though differently, in Idaho, New York, New Jersey, Georgia, and Indiana

Other examples include the body of FTC Safeguards Rule's broad requirement for Due Care when handling data, as well as the Federal Information Systems Management Act (FISMA), a recent bill filed (then withdrawn) by Rep. Putnam of Florida, US Senate S1350 (Cal SB1386 clone), and the Computer and Information Security Working Group (CISQG) are not even the entire list.

Metrics: not free

- Process-based metrics work only in stable attack environments
  - Damp change to get stability?
- Goal-based metrics will have to be indirect because security is a means, not an end
  - Is ordinal scale good enough?

Metrics are not free. More particularly, there is the natural tendency of government and wanna-be governments (like the insurance industry) looking to impose process standards. Process standards are only valuable against constant threat, what an academic would call a stationarity assumption. This does not obtain in the Internet sphere where rapid technical advance is a desirable economic good. Therefore, the national policy level question is whether damping down the rate of change to a more stable state, one treatable by process standards, is a desirable direction for the national leadership to take.

By contrast, goal-based metrics at our present level of knowledge will have to be indirect as we have no way to measure security, per se. What we can measure are downstream effects of security or, more precisely, downstream effects of insecurity on, say, unreliability, as illustrated in the next slide. What we can measure is likely to be at best an ordinal scale but isn't an ordinal scale good enough? If not, say why not.

Ends *v.* Means

If a system is insecure, then
It is unreliable, therefore
Security is necessary for reliability, yet
Security is insufficient for reliability, ergo
Security is a subset of reliability.

The more mature tñnfrastructural entity is the more security is a subset of reliability, per the logic above.

The parallel: that if a system is unregulated then it is unpredictable, therefore regulation necessary is for predictability, yet regulation is insufficient for predictability, therefore regulation is a subset of predictability suggests itself. If as correct as the relation between security and reliability, then the question for the law is how to regulate for predictability without damping out innovation or the motivation to improve. This is hardly a new topic, but the digital physics will stress security as a subset of reliability.

As Whit Diffie (Stanford) has observed, computing would become free were it not for security.

Security spend

"The next ten years will be a referendum on whether we consume the entire productivity growth of the US economy for increased security spend." [*paraphrase summary*]

Chief US Economist, Morgan Stanley
Op-Ed, NY Times, 23 October 2001

"The Terror Economy," Richard Berner, NY Times, 23 October 01, Page A23, Column 1

ABSTRACT - Op-Ed article by Morgan Stanley economist Richard Berner warns that war against terrorism will impose long-term economic costs in form of higher insurance and security costs, maintenance of larger inventories and new Internet security measures; explains that spending more on defense will erase decade-long 'peace dividend' and crowd out other investments that helped transform budget deficits into surpluses.

Full article available upon reasonable request.

Security spend as calibrator

- Corp budget for security:
  3% for manufacturing...8% for banks
- IT headcount for security:
  5% of total
- IT budget for security:
  12% hardware 20% software
  15% services 53% staff

source: Meta #2856

The Meta Group, Diamond report #2856, recommendations on how much of IT budet should be allocated to security spend.

The problem of value

"Some day, on the corporate balance sheet, there will be an entry which reads, 'Information'; for in most cases the information is more valuable than the hardware which processes it."

Grace Murray Hopper, USN (Ret)

Rear Admiral Grace Murray Hopper, USN (Ret), Washington, D.C., 1987.

The problem of value

- How much is information worth?
  - Replacement value
  - Black economy market price
  - Future value

How much is information worth, then? If Hopper is right, then it ought to be on corporate, and for that matter national, balance sheets and it generally is not.

Perhaps we need a way (or ways) to think about the value of information. Perhaps its replacement value, or its value in the black economy, or its future value would at least bracket reality.

The problem of value

- Replacement value
  - How much would it cost to build a brand as good as the one you have today?
  - What is the time to recycle after a continuity break?
  - Management cost of new passwords for 5,000 users

You ask a management team "How much is your brand worth?" and you get blank stares or wild guesses. Try it a different way, ask "How much would it cost you, knowing what you know today, to build a brand from scratch as good as the one you have now?" This will get an answer that is probably a lower bound for replacement value. If such a value is sufficient basis to make whatever managerial decision around security that is on the table, then that is good enough for the time being.

Similarly, if your business has a "non-interruptibility" requirement, such as continuous monitoring of weather conditions for a period of time before a power plant can be sited, then the re-formulated form of "How much is your information worth?" would be more like "How much incremental cost would you incur if your continuity of measurement were broken and you had to start over?"

A different sense of the value of good passwords or good password protection would be to not ask "How much are your passwords worth?" but rather "If today you had to get all 50,000 people in your firm to pick a new password within 36 hours how much incremental cost would you incur?"

The problem of value

- Black economy market price
  - AOL screen names: 0.1¢/name
  - Bot-net host rental for spam: $1/mo
  - Financial screenshot: $500
  - Game skin 90 days out: $50,000

A different way to look at the value of information is to ask what the black market pays, if indeed that is a question that can be answered in a way that is sufficiently close to where you are to be valuable via analogy. For example, a thief was paid $100,000 for 92MM AOL screen names.

Computers that are taken over silently are occasionally rented to others, e.g., as spam relays. The rental fee approximates $1/month by some estimates. That tells you at the very least that the supply of machines taken over is great as such a price is obviously slight. That would mean that your data on your machine is, by analogy, very easy to get at by others. If you don't know how easy it is, then you would conservatively assume that breaking into your machine is worth a dollar on the open market.

More directly, a major west coast bank reports that its tellers are routinely offered $500 per screenshot of customer identifying data for customers with over $50,000 of assets. So a clerk making $10/hour can give themself an after-tax raise of $26,000/year for the price of one sheet of paper per week. Not every clerk is immune to this temptation.

Game skins more than ninety days pre-release are worth at least $50,000 in Taiwan.

The problem of value

- Future value
  - From eureka to FDA filing costs circa $100M, 80% is information
  - Derivative pricing algorithm alone carried on books as $300M
  - Patent losses: CDMA in India & China at $750M/annum

In a pharmaceutical company, the critical period begins with the "Eureka!" moment and closes with the FDA formal filing. In this interval, the pharmaceutical can expect to spend $100,000,000 at the end of which 80% of the value is the information in the can. This is a hard to get figure and was obtained in conversations variously.

A single bank in NYC that is known for its derivative trading carries its apparatus for pricing same as a $300,000,000 asset.

The inability of Qualcomm to effectively patent its CDMA technology in China and India represents an information loss to them of $750,000,000 per year based on current usage rates of the CDMA technology.

Pre-emption

"What did he know and when did he know it?"

<issue, Congressman, date>

Every single pointed, argumentative, accusatory discussion in the press, the salons, and the hearing rooms of Congress comes down to the phrase "What did he know and when did he know it?"

The subtext is clear, the hostile question will immediately be followed with "He should have know and taken action earlier than he did. He should have kept XYZ from happening." In short, he should have pre-empted.

Pre-emption

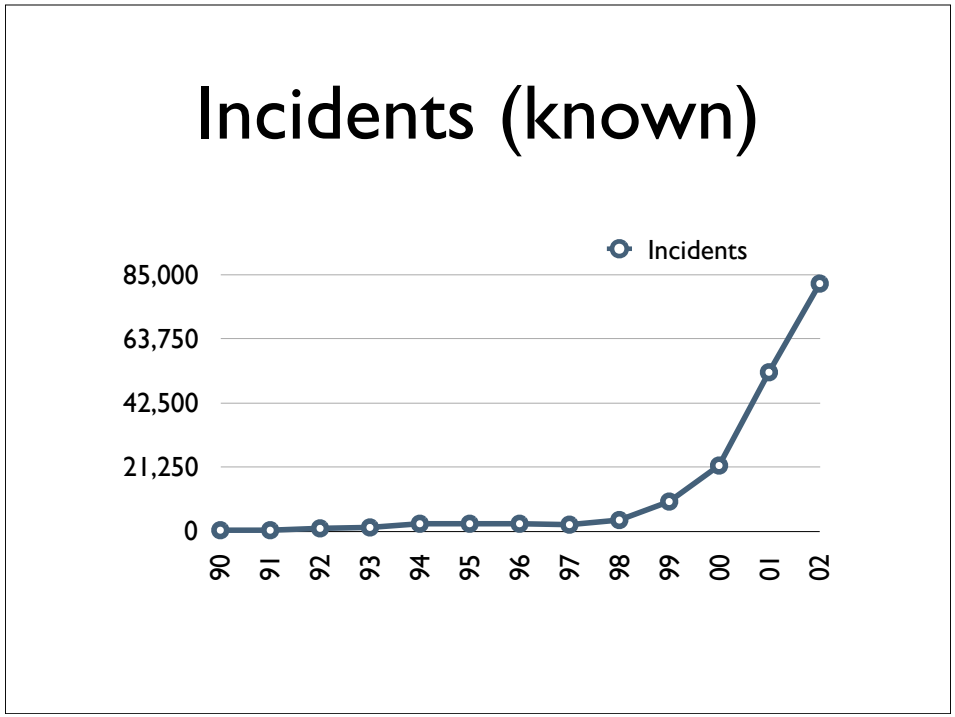
- Invisible foes create demand for pre-emption
- Pre-emption requires intell which requires surveillance
- Surveillance requires a sensor fabric that is always on

Freedom (default permit) yields to Safety (default deny)

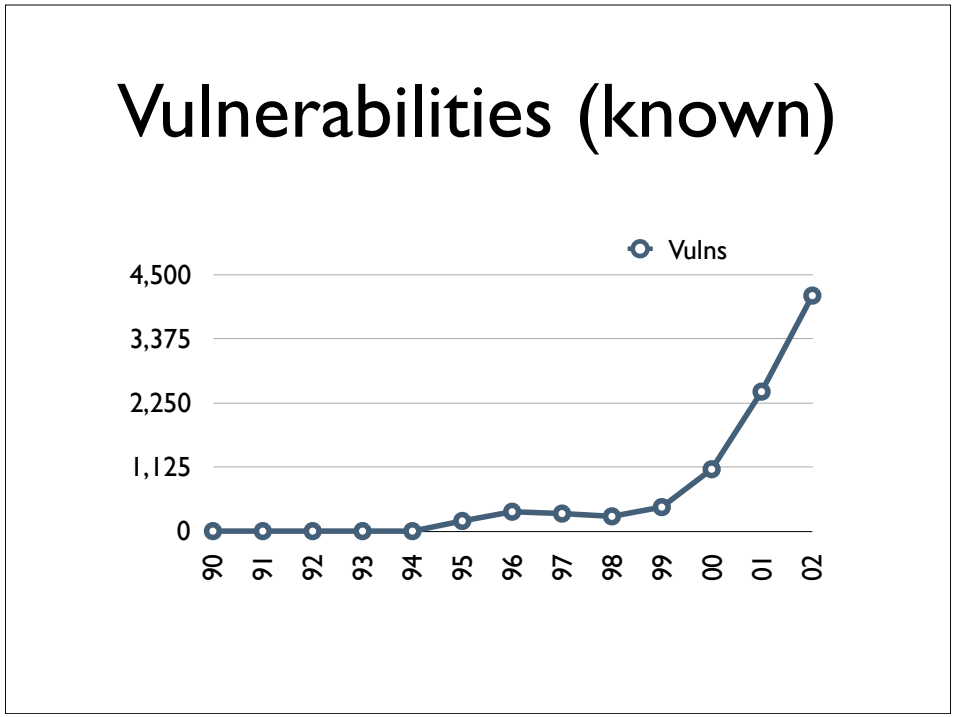
Here are the facts of pre-emption.

Invisible foes create an unblockable demand for pre-emption of what those foes would otherwise do. For pre-emption to work, there must be intelligence on what might happen, and when and where and how as well. For that intell to be in hand, there must be surveillance before there is any proven reason to be surveilling, i.e., it must be done when suspicion is the most the surveiller can have. In a highly interconnected, globalized, fast-paced and fast-changing world this means a sensor fabric that is always on. The first rule of exploratory data analysis is "Get the data." As a matter of national security, there must be significant spend on surveillance and that spend may not be only in dollars but rather in social costs as well.

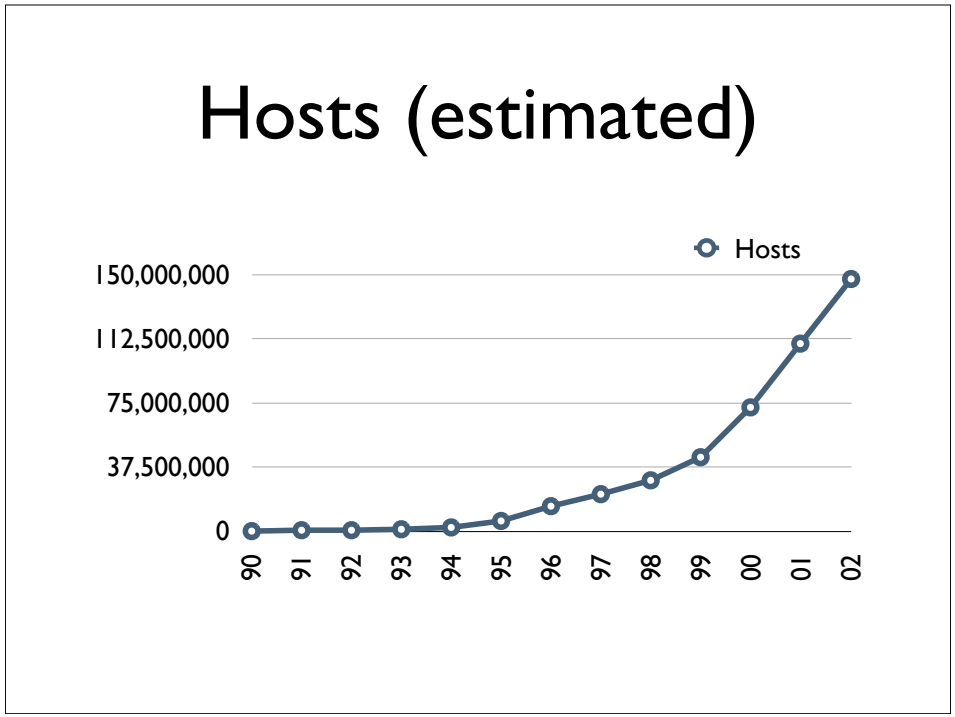
It is time to illustrate these points.



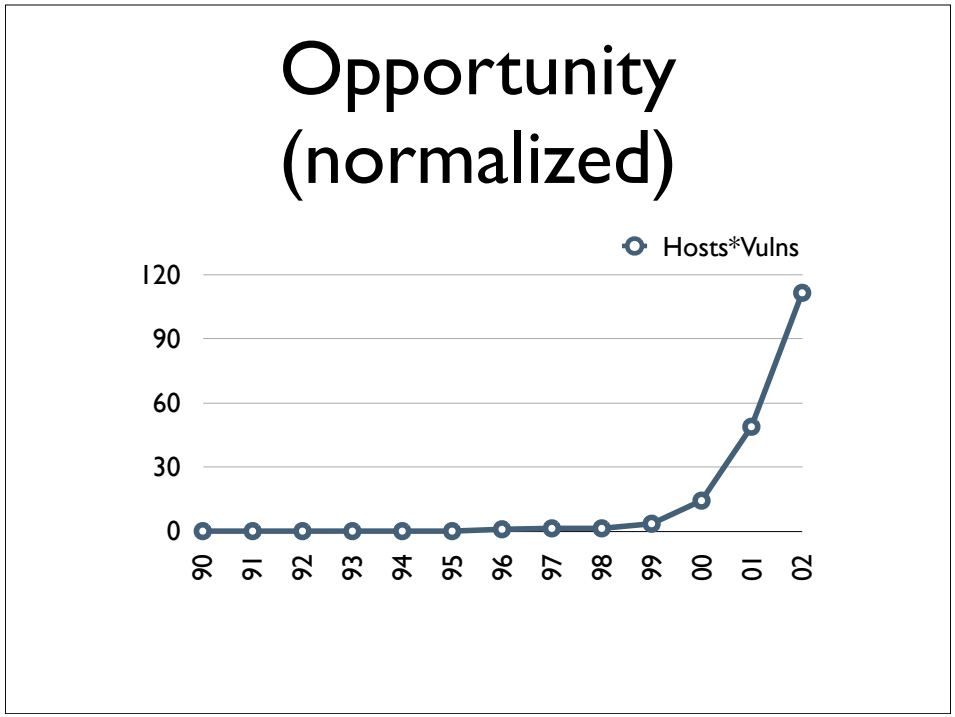
Public CERT data



Public CERT data

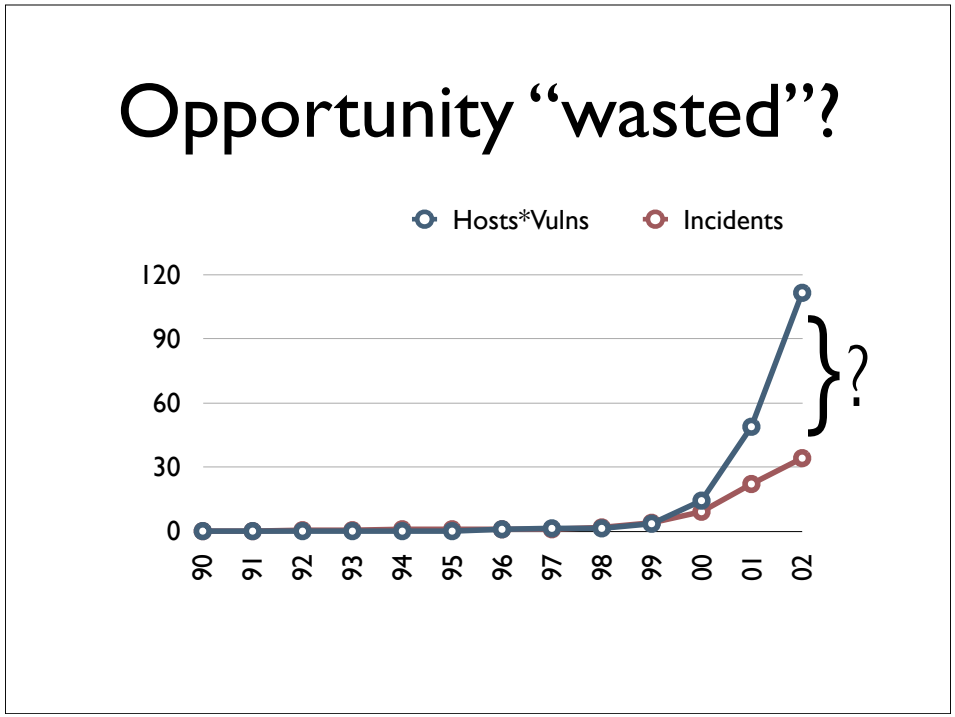


Public ISOC data



So how much opportunity is there? Is it well modeled by total number of open holes, i.e., by the product of the number of hosts times the number of vulnerabilities?

If so, the curve looks like this, and it has taken an amazingly steep turn upward.



If opportunity is proportional to the product of hosts times vulns,...

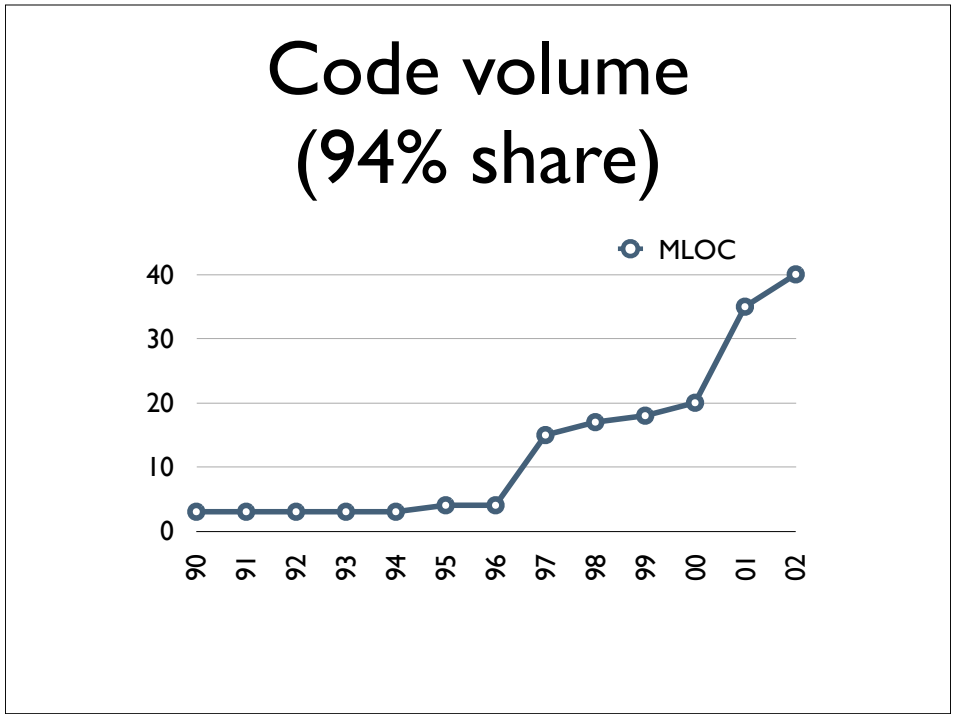
Then either "we" are doing a good job at keeping the crime rate from growing as fast as the opportunity is growing, there is some degree of the attack community holding vulnerabilities in reserve, or there is a growing reservoir of untapped opportunity for attack.

Complexity

"There are two ways of constructing a software design. One way is to make it so simple that there are obviously no deficiencies and the other is to make it so complicated that there are no obvious deficiencies."

C.A.R. Hoare

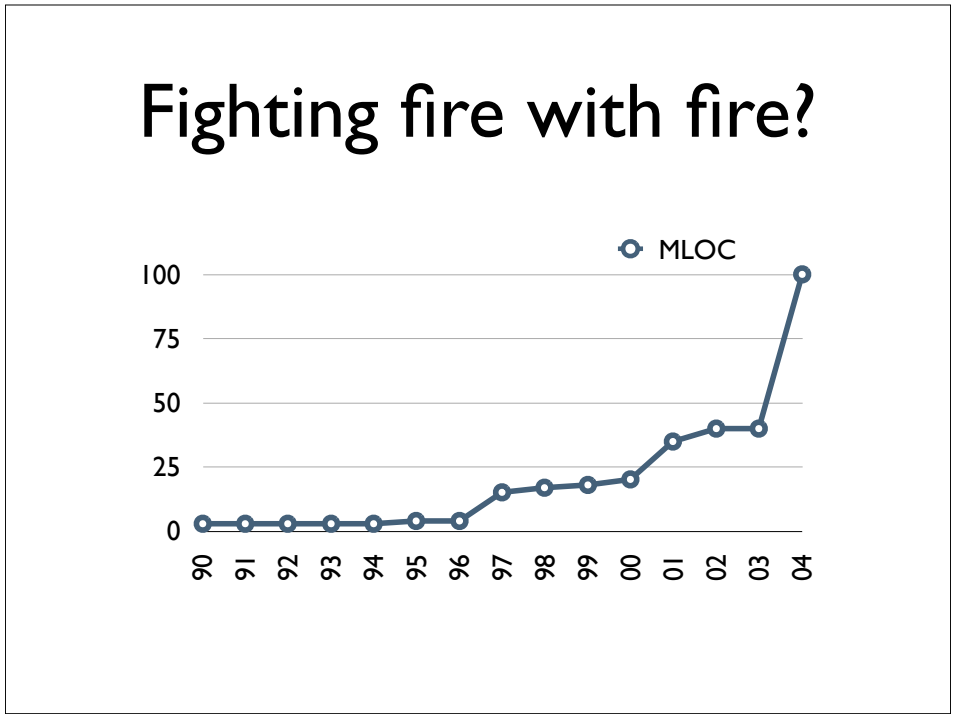
This sums up the question of complexity. The parallels to current market leading suppliers, competing as they are on feature richness, is obvious and daunting.



Windows 94% market share per IDC

Code volume as observed:

Win 3.1	Win NT	Win 95	NT 4.0	Win 98	NT 5.0	Win 2K	Win XP
3	4	15	17	18	20	35	40
1990	1995	1997	1998	1999	2000	2001	2002

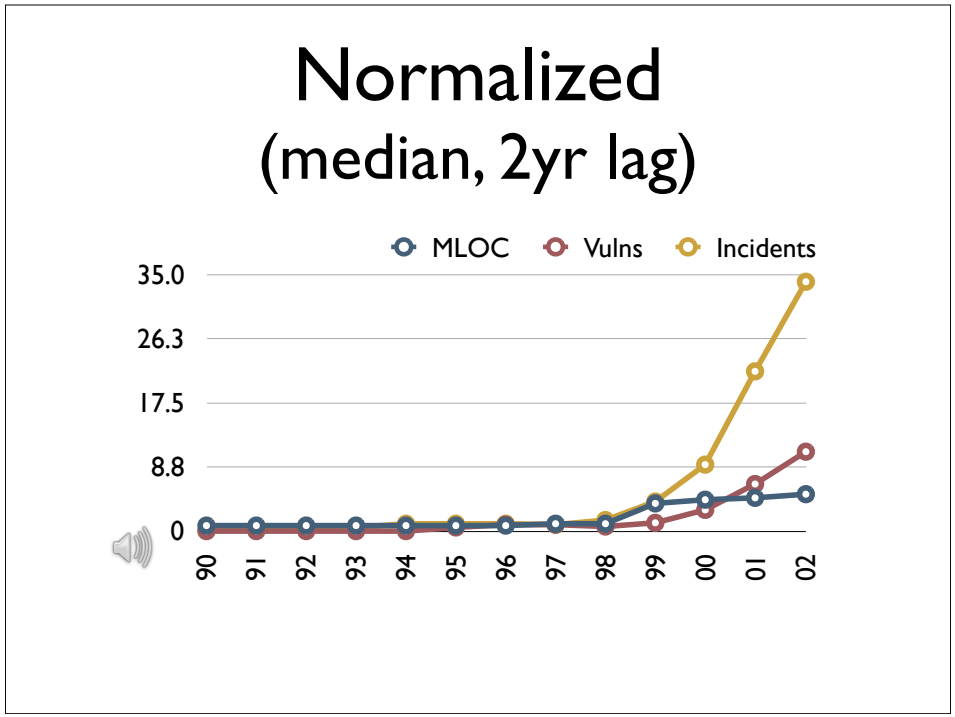


Code volume as observed:

Win 3.1	Win NT	Win 95	NT 4.0	Win 98	NT 5.0	Win 2K	Win XP	Longhorn?
3	4	15	17	18	20	35	40	100?
1990	1995	1997	1998	1999	2000	2001	2002	2004?

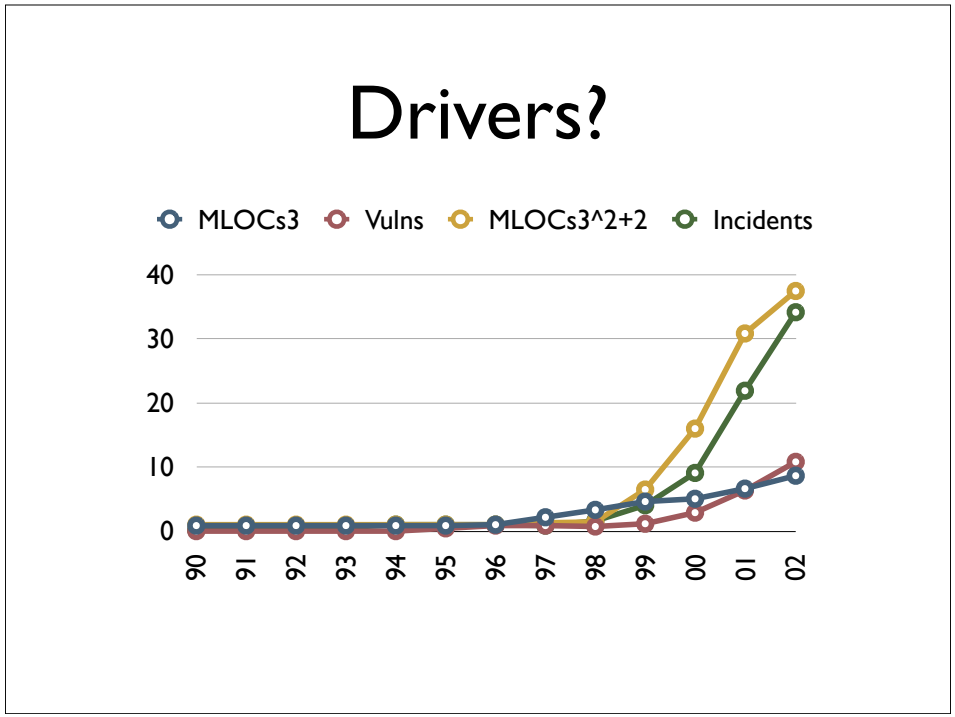
How big will Longhorn be?

For an historical comparison, look at the testimony of David Parnas on the implications of President Reagan's "Star Wars" anti-ballistic missile system proposal. He resigned from the study commission on the grounds that 100 MLOC was an utterly preposterous thing to imagine working well enough to rely on it for critical things like national defense. Herb Line, then of the Center for Strategic Studies, thought 10 MLOC was preposterous. And here we are.



Each curve is normalized against its own median over this period. Therefore, overlaying the curves is legitimate.

Code volume curve is shifted right two years to crudely simulate diffusion delay.



Each curve is normalized against its own median over this period.

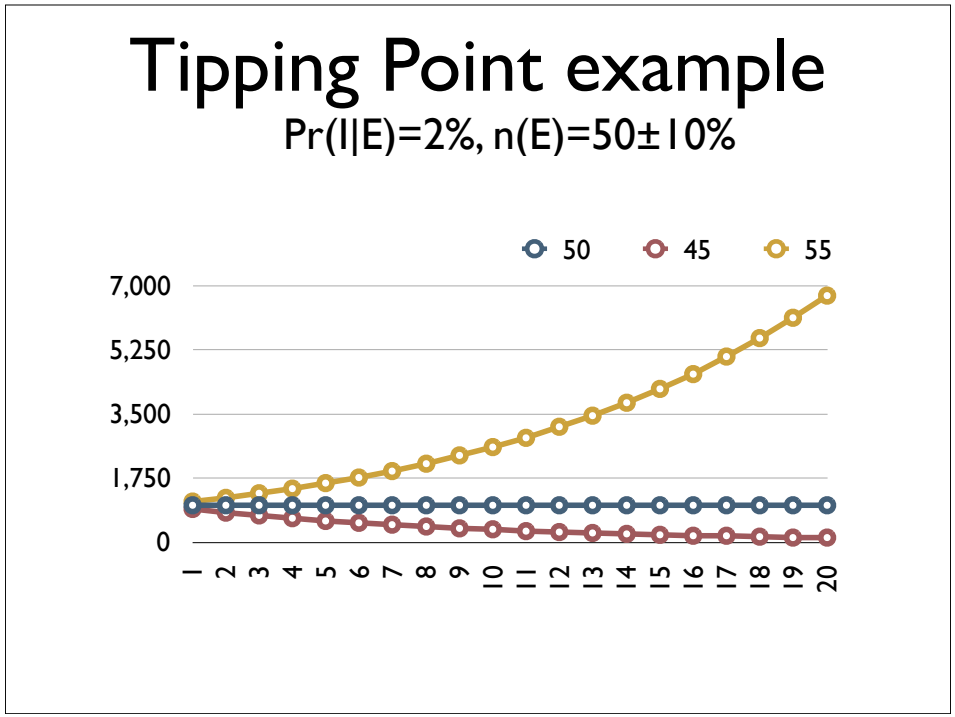
Code volume curve, MLOCs3, is the three year moving average of code volume, perhaps a better estimator of effective code volume in the population at large.

The second code volume curve, MLOCs3^2+2, is the square of the three year moving average of code volume, and then shifted right two years. The argument is this: Security faults are a subset of quality faults and the literature says that quality faults will tend to be a function of code complexity, itself proportional to the square of code volume. As such, the average complexity in the field should be a predictor of the attack-ability in an a priori sense. Shifting it right two years is to permit the attack community time to acquire access and skill to that growing code base complexity. This is not a statement of proven causality -- it is exploratory data analysis.

Epidemics

- Characteristics of infectious processes
  - Pr(infection|exposure)
  - interval from infection to infectiousness
  - duration of infectiousness
  - interval from infection to symptoms
  - duration of acquired immunity

The math for modeling epidemics is well developed, as is the math for accelerated failure time testing, actuarial science, portfolio management, and others. There is no need, and no time, to invent new science before progress can be made. Steal these skills, and do so while the senior practitioners in security still include people with these sort of skills learned elsewhere.



This is simply the example used in Gladwell's The Tipping Point. It illustrates the chaotic nature of epidemics which is to say that small changes in initial conditions produce large changes in downstream values. This example is where the initial number of cases is 1,000, the probability of infection given exposure is 2%, the number of exposure events while infectious is 50 plus or minus 5 (10%), and the downstream shows that in only 20 days at -10% the disease will die out while in only 20 days at +10% the epidemic will be well underway.

Worst case disease

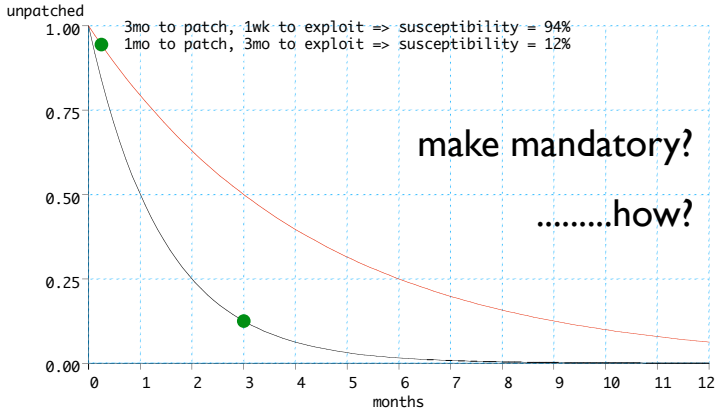
- Pr(infection|exposure) = 1.0
- interval from infection to infectiousness = 0
- interval of infectiousness = open ended
- interval from infection to symptoms = indef
- duration of acquired immunity = 0 (mutates)

If you were designing a pessimal disease, it would be perfectly transmissable (100% chance of getting the disease once exposed and no acquired immunity), no symptomatic sign of infection, and an instantaneous conversion from pre-infection to infectious (or from prey to predator, if you prefer).

The above describes worm propagation, or DDOS zombies, or the stockpiling of unannounced vulnerabilities.

Does the law have an answer for designer disease with pessimal characteristics and self-obscured authors? Is “terrorism” an appropriate model or is it more like mandatory seat belt laws?

Immunization



Qualys, Inc., has data that implies patching is like radioactive decay in that 50% of the remaining unpatched systems will be patched in each succeeding “half-life.” Qualys’s figure is 30 days.

Posting a patch starts a race wherein the patch is reverse-engineered to produce exploits. The two data points are intended to bracket current reality. In the one case, if patching does have a one-month half-life while the reverse engineering interval is 90 days, then the susceptibility would be 12% at the moment of exploit. By contrast, if patching has a three-month half-life while the reverse engineering interval is one week, then the susceptibility would be 94% at the moment of exploit.

Time-to-exploit is shrinking while the time-to-patch is lengthening (if you factor in the growth of always-on, always-connected home machines) so the question becomes whether “mandatory” is a word we must use and, if so, what would it mean. What does the law say?

Durability tradeoffs

- Durable against random faults
  - Scale-free networks (growth driven)
- Durable against targeted faults
  - Structured routing (policy driven)

...cannot be both

Research, practice, and history each point to the same conclusion: Those network structures that are optimized for resistance to random faults are not the networks that are optimized for resistance to targeted faults. This is not a happy tradeoff.

Networks tend to grow by accretion and new nodes will prefer to be connect to nodes that are well connected, a phenomenon that produces so-called “scale-free” networks. These networks are remarkably resistant to random faults, and the Internet is to a large degree characterizable as a scale-free network. However, because of the path of any given packet will tend to pass from lightly interconnected nodes through highly connected nodes, this resilience to random faults also makes the network vulnerable to targeted faults. A body of scientific literature is growing up around this, beginning at <http://citeseer.ist.psu.edu/407844.html>.

Side issues abound

- Tight integration of apps & OS
- User level lock-in
- Decreasing skill/power ratios everywhere
- Insecure complexity *v.* complex insecurity
- Strength through diversity
- Opened source *v.* open source

This list is indicative, not exhaustive. It includes the monopolization questions of tying the applications to the operating system thus to insure that a security failure of the one is a security failure of the other, whether user level lock-in plays a role in assessing the locus of liability for security faults, and whether the skill to operate ever more powerfully interconnected computers does not at some point require some a priori proof of capability. It asks to distinguish complexity that is insecure from insecurity that is itself complex. It ponders the question of genetic diversity as a survival advantage in a world where predators have just arisen. It distinguishes the value of public disclosure in the open source tradition to the private disclosure of the entirety of the Windows (94% share) source code pool to potentially hostile nation states. It could go on. The challenge is substantial and historically crucial. What will the law say, and can it say it without adding noise?

Exploration

- Latency (to patch, to detect, MTBF, MTTR)
- Interarrival rates (attacks, patches, unknown hosts)
- Intrusion tolerance (diversity *v.* redundancy)
- Comparands (benchmarks, shared pools, anova)
- Cost effectiveness (risk reduction *v.* symptom relief)
- Scope (data capture *v.* data reduction, sampling)

To go on from here we can't use words, they don't say enough. We must use numbers. These are indicative and intended to push you to think of more. Even if the shorthand does not read clearly, the point is this: now that the digital world is essential, statistics based on the realities of digital physics will be, at least, how score is kept. Perhaps we will be fortunate and statistics based on the realities of digital physics will also inform decision making at the highest levels, including the law.

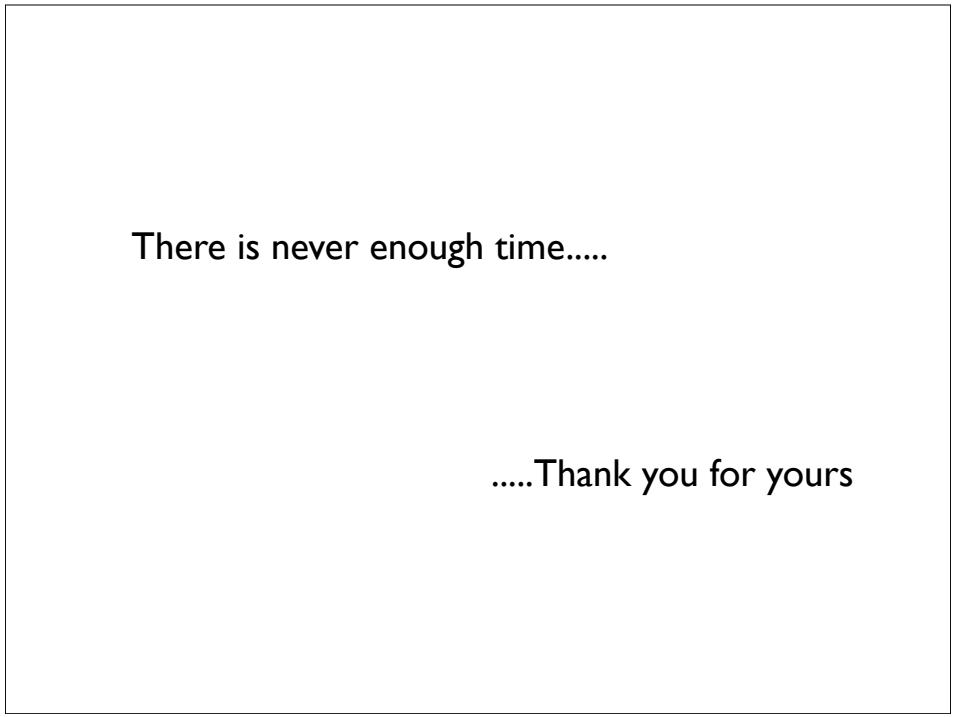
Summary

- Unknown vulns = secret weapons
- Absence of events does not predict calm
- Mobile-code mandates trade downside risks against each other
- Risk is proportional to reliance when the relied-upon cannot be measured
- Price of freedom is the probability of crime

In summary, the pool of selectively known vulnerabilities is the secret weapon of the serious enemy, the absence of a significant catastrophe to date is most assuredly not evidence that the risk is low because in a risk aggregated world significant events make up in their severity what they lack in frequency, that mandates for automatic patching are effectively mandates for more powerful mobile code and are thus risk creating in the larger sense of risk aggregating, that risk is itself proportional to the reliance on places in the entity being relied upon exactly when there are no effective measures, and that the tradeoff between freedom (default permit) and safety (default deny) is real and present.

A modest proposal

This is the last time we will have as much hybrid vigor amongst leadership as we do now and the last time we will have as clean a slate as we now have; we must use them both for all they are worth.



It has been entirely my pleasure.

Dan Geer
dan@geer.org
+1.617.492.6814

challenging work sought & preferred

Further contact is welcome particularly if it brings problems of the sort that illustrate the bounds of our knowledge.