The Physics of Digital Law: Searching for Counterintuitive Analogies

Daniel E. Geer, Jr.

"Digital law" is and must be counterintuitive – an intuitive understanding of sticks and stones does not translate to digital worlds. Because our intuition about the digital sphere can so easily be wrong we need to substitute solid facts for faulty intuition. It is said that the practice of law is a search for analogies, so law is most susceptible to mistakes when the digital reality differs from our commonsense intuitions about the physical world. In this paper, I use the neologism – "digital physics" – to describe the important features of digital spaces and the parameters they set on computer security. Some of the principles of digital physics are directly related to concepts of physics like space and time. Others have no physical world analogy, but rather describe a fundamental truth about the digital world in the same way that the laws of physics describe fundamental truths about the physical world. Just as these laws of physics dictate what is and is not possible in the physical world, the laws of digital physics dictate what is and is not possible in the digital world, which will prove crucial for making policy choices.

In this chapter, I offer a guided tour of some of the important problems of computer security, including risk management and the interaction between trust and identity. These examples not only show how our intuitions about the physical world can be misleading in the digital world, they also provide a good model for approaching the other issues addressed in this book.

A. Risk and Risk Management

Risk is not an enemy of good policy, per se, but unmanaged risk can be. Sometimes, risk is easy to recognize – for example, climbing a high tension transmission tower in a thunderstorm. When it comes to computer security, however, risk is harder to understand, harder to apportion, harder to manage, and harder to clean up after.

One reason why digital risk is hard to manage is the lack of warning time before an attack. In the physical world, the bigger the attack the bigger the warning time – for example, you can see an invading army coming from a long way off. Our intuition about time in the physical world might lead us to believe that we will be able to see a cyber-attack coming and have time to defend against it. But a principle of digital physics challenges this intuition: information travels faster over the network than humans can react to. Thus, in the digital world the warning interval between detection and attack decreases towards zero. And after that brief interval between detection and attack, further attacks often materialize with no warning at all.

Not surprisingly, infections become more effective over time. For example, the infection rate of the infamous Slammerworm doubled every 8.5 seconds.[1] The Witty attack on various ISS firewall-related products offered only two days warning between announcement of the vulnerability and the appearance of an exploit of it in the wild, any further contraction of warning interval is operationally irrelevant.[23] As this was written, the Sober.P virus has gone from first detection to 77% of all inbound virus payloads in email – in a matter of sixty hours. It is delusional to believe that we will have sufficient reaction time to prevent cyber attacks.

These are examples of "cascade failures"--security failures where the infection converts prey to new predators. The only way to control them is to limit the rate at which the failure spreads. Much like a dangerous crown fire moving quickly upward on steep terrain, spread rates for digital infections are too fast for reaction on a human time-scale. As a result, the only way to protect the Internet commons is, as with forest fires, to block further spread by pre-cut "firebreaks."[4]

In the network world, preventing the outbound spread of toxic digital traffic is called "egress filtering"; preventing toxic digital traffic from flowing inbound is called "ingress filtering." Should the law require egress filtering as a sort of digital firebreak, or should it instead try to block incoming toxic content? Many enterprises, like universities, which have strong free speech traditions, have policies against blocking suspect inbound communications, while others enterprises are simply too poorly managed for effective ingress filtering. Thus, egress filtering may be the Internet's last line of defense. It is helpful to make an analogy with public health law: While it is not criminal to get a disease, it may be criminal to break quarantine and pass the disease on.

Thus the law might make enterprises pick from a menu of options including ingress filtering, egress filtering and immunizing its computers with regular security patches. To be effective, enterprises will be forced to choose at least two.[5] But of these options, egress filtering may be the most important. A bank will certainly want to protect its internal systems from incoming attacks but it may be even more important for the law to require that the bank's computers not be the source of an attack on others.

A machine that has already been penetrated by a successful attack (hacker slang is "0wned") may exhibit no outward symptoms. This defies our physical world intuition that as long as a machine is doing what it is supposed to do, it is probably not broken. Digital physics teaches us, however, that the outward appearance of a "working" machine will not tell us whether the machine is doing *more* than what it is supposed to do. Hence the user might never have any reason to detect that the machine is under an attacker's remote control; in this way the culprit can stockpile 0wned machines.

There are technical measures to detect 0wned machines. Even so, we must decide who will be responsible for the scanning required to detect infection. Should the responsibility fall on the owner of the machine, or should it fall on the owner of the transmission capacity—for example the Internet Service Provider (ISP)—that an infected machine uses when it tries to attack other sites?

Should we require ISPs to scan their customers' computers? ISPs are certainly capable of doing so, but scanning a host machine for vulnerabilities will itself trigger warnings on the scanned host. If the ISP learns that Machine 12345 is infected, should the ISP then be obligated to notify, to terminate service, to report to other authorities, or to put that host under increased surveillance?

These questions might focus on the wrong level of concern. Perhaps the proper analogy is public health: It was public hygiene and not individual medical treatment that solved London's cholera epidemics in the nineteenth century.[6] Today, in the United States, the Centers for Disease Control play a global role, which has become ever more essential as the world has grown smaller. The CDC regulations include mandatory reporting, forced treatment and/or quarantine of the ill, publication of an authoritative Morbidity and Mortality Weekly Report, formal predictive work to aid officials who seek to anticipate epidemics, and public identification of locales with an excess of incidence or prevalence of any particular pathogen.

In a 2002 paper, Stuart Staniford and his collaborators proposed a CDC for the Internet[7] and in the fall of 2004 the NSF funded this as the Center for Internet Epidemiology.[8] How closely should the CIE parallel the CDC? A close parallel would require extraordinary things –mandatory reporting of security incidents, forced quarantine of sources of electronic infection, publication of definitive analysis of infection sources and vectors, and the kind of longitudinal expertise that identifies countermeasures before crises occur. No corporate counsel will happily accept mandatory reporting. Suppose such reporting requires entities to turn over logs from their Internet-facing networks. These logs may contain evidence of an attack the entity failed to notice, and as such, the logs may provide evidence of the entity's negligence. On the other hand, if an entity does not share its network logs, it is impossible to determine if it is a specific *target* of attack, or if it was just opportunistically attacked along with many other random targets. Despite the long term benefits of sharing logs, most corporate counsel, who are paid to be risk adverse, will resist both voluntarily sharing and regulatory measures that mandate sharing.

The few attempts at voluntary sharing of information have had very limited success. The Information Sharing and Analysis Centers, created by President Clinton in 1998,[9] are "sector-specific," meaning that information gathered by entities in the financial services industry are not necessarily shared with information gathered by the telecommunications or information technology industries. In an attempt to assuage the fears of corporate counsel the data collected by the ISACs is exempt from the Freedom of Information Act and from antitrust laws. These exemptions, however, have not yet been tested in the courts, and because those exemptions are untested, information sharing has been modest at best. All of this may change however, if the Department of Homeland Security's Infrastructure Analysis and Protection Authority succeeds in its bid to take over the ISACs.

A central issue is whether and how the private sector of the US economy, which owns 90+% of the nation's critical infrastructure, can ensure that infrastructure's protection from shared risks. Government's interests and industry's interests align when stated abstractly but less so at the levels where details matter. In some cases Federal regulation may be necessary while at other times self-regulation works just as well. Generally, lawyers are paid to be risk averse and to act in the particular interests of their clients. In the digital world, however, private tragedy can produce a public debacle. Self-regulatory solutions require incentives for private parties to share information, and this may require immunizing companies from liability. Government regulation, on the other hand, must balance protection against privacy. Neither solution is an easy task.

Thus far we have been discussing the risk of cascade failure, which begins with a single computer or a small number of computers. There are, however, many risks on a national scale: the collapse of an essential component of the public Internet, the discovery of infiltration across many firms, the leak of an attack method against which there is no current defense, and a loss of public confidence in the banking system as a whole after a security failure.

How do we assess the risk of one of these failures occurring? If there has been no overt catastrophe to date, can we conclude that risk is low? Once again our intuitions about the physical world fail us. It does not follow that the absence of a history of loss means that the risk is low. The key question is whether the events that produce loss arrive more or less independently, or whether they tend to come in bunches. If the risks are small and spread out at random you would expect to see some events over any sustained period of observation; if you do not, then that means that the risk is quite low. If, however, the small events come in bunches, then it may merely be that their time has not yet come.[10] If attacks in the digital world are mostly independent of each other, then recent history does have some (limited) predictive power; but if the attacks are calculated, rather than opportunistic, then a history free from a large scale attack is not particularly meaningful.

The question for the law is what we should require of the owners of digital assets, especially those assets that are part of the critical information infrastructure. Do we require that these owners plan for disaster in order to mitigate it when it comes? Do we require them to share their plans with coordinating agencies? If not, how do we ensure their preparedness?

The problem is that each attack always requires two responses: First, we must regain operational recovery; second, we must preserve evidence of the crime. Without an effective plan for handling incidents, these goals tend to diverge – you can have prompt operational recovery at the cost of losing forensic evidence or you can cleanly recover forensic evidence but at the cost of delaying operational recovery. Should we impose reliability standards for digital assets in the way we already do for electricity generation and clearance of banking transactions? Or would it be better to require incident reports whenever intentional damage cannot be ruled out? Either way, we must alter our physical world intuitions about the relationship between private and public risk. The idea that “It’s my computer and I’ll protect it if want to” makes no sense in a digital world; every important cyber attack has involved the recruitment-- or, more correctly, the hijacking-- of the computers of unsuspecting users. Digital physics teaches that your failure to protect yourself is not just a risk to you; it is also a risk to me.

We assume that the government's job is to protect us from those things that we cannot protect ourselves from. But if private tragedies in the digital world threaten public security, we cannot separate purely private risks from national ones. If so, is there a risk management strategy at the national level that can protect against national tragedies? Digital physics can assist us by delimiting the possible classes of security failure. Only two classes of failures qualify as threats on a national scale: 1) inherently single points of failure; and 2) cascade failure.

A "single point of failure" is a part of a system which, if it fails, interrupts the entire system. Inherently single points of failure are those which by necessity or design must be so. An example is the fabled Red Telephone that the President uses to communicate with key foreign leaders; nine Red Telephones in the President's office would be worse than one.

In the digital world, the assignment of network addresses is a single point of failure because without a unique authoritative source of network information there would be chaos. If instead of a single Domain Name Service (DNS) there were multiple sources of network information, different users could receive different answers to their requests to be directed to a particular URL. However, the Domain Name Service must not only be a single authoritative source; it must also be reliable. In reality, responses to queries may come from any of thirteen primary service providers which provide identical answers. Disabling them all simultaneously is hard; we know that because it has been tried before. This sort of redundancy is essential for an inherently single point of failure, and it is yet another example of where our physical world intuitions do not work well in the digital world. In the physical world we do not expect that a single authority, say, the King of England, is a position that could be held simultaneously by thirteen different individuals each of whom is guaranteed to make the same decision regardless of which one is asked--and you won't be able to tell which one is answering anyhow.

The classic method for protecting a single point of failure is "defense in depth" – that is, using rings of protection where each ring of protection exists to catch the failure of the ring outside of it. No single ring of defense need be a technical masterpiece nor is any particular ring—considered in isolation—difficult to get through. Rather, it is the sum of the rings that secures the protected object. Medieval castles relied on defense in depth, as do modern prisons (though in this case the rings serve to keep the prisoners in).

Defense in depth works in the digital world as well. For example, the root servers that operate the Domain Name System are protected with (1) redundancy (so that an attack on any of the thirteen servers does not affect the other twelve), (2) caching (creating a short term memory on individual computers which makes them less dependent on the root servers), and (3) diversity in design (none of the root servers are designed the same way, so that a successful attack on one server does not increase the risk of a successful attack on the other twelve).

If we wanted to, we could add more root servers or require ISPs to put upper bounds on how many nameservice requests they could forward per second. In other words, we could add more levels of defense. In this sense, defense in depth is not a technical problem, but rather a matter of how much money we want to put into the project.

Our real attention, then, should be on the second class of failure at the national scale: cascade failure. In a cascade each part that fails increases the probability that the next part will fail too. A row of dominoes or a multi-car pile up are familiar examples. Infectious diseases are also cascade failures. Because of international air travel, a break-out of the flu in Hong Kong increases your chance of getting the flu in New York; any place where people congregate in large numbers, such as airports, increases the transmission rate.

Digital viruses are cascade failures, and the main contributor to the problem they pose is digital monoculture—the fact that most people use the same operating system or other key software components. The term "monoculture" reminds us all that healthy ecosystems are diverse. Unlike the defense of singular points of failure, there are no easy solutions to protecting against cascade failures. Defending a singular point of failure assumes that there are insiders and outsiders. A good system of defense-in-depth prevents outsiders from getting in, leaving only the more difficult problem of preventing insider attacks. But in a software monoculture, no one is an "outsider;" to the contrary, everyone is an "insider." When everyone uses the same software, an effective attack against one computer can be an effective attack against all. And if that attack can be automated, then only network bandwidth constrains the spread rate.

Again, this defies our intuitions about the physical world and the spread of infectious disease. You cannot kiss me or sneeze on me from an arbitrary location anywhere in the world, and you cannot do so without revealing that it is you; but you can do both things in the digital world. Distance, as a concept of digital physics, is radically different than distance in the physical world. In the digital world, every computer is essentially equally distant and equally close to every other computer; this makes the risk of monoculture far more frightening than our intuitions about the physical world suggest. To understand this, one need only look at the spread-rate videos of computer viruses that the Cooperative Association for Internet Data Analysis (CAIDA) puts out; what we see is a simultaneous worldwide crescendo, not ripples in a pond spreading neatly from the point the rock entered the water. [11]

What should the law do about the shared risk to the national economy directly traceable to our common digital vulnerabilities? We refuse children admittance to public schools unless their parents agree to immunize them against common biological vulnerabilities. We grant special status to public utilities because of their status as natural monopolies but regulate the quality of their services because that monopoly creates a common vulnerability. In the digital economy, however, we have yet to face the shared vulnerability produced by our digital monoculture; indeed, even the Department of Homeland Security acknowledged on the floor of the US House of Representatives that it is internally a computing monoculture and that this practice creates risk.[12]

B. Identity and Trust

In addition to the over-arching issue of risk and risk management, we must deal with questions of identity and trust. The two issues are conjoined, and in the digital world our intuition and the digital reality tend to diverge.

In a digital economy, information is the coin of the realm. Electronic commerce requires amassing useful information and putting it to productive use; the challenge is keeping that information valid and available only where it should be. Getting the most use out of the business' information while not losing control of it is the core of a successful electronic business. For example, the information on Lexis/Nexis would be useless if people could not search for and find it, but equally useless if people could not trust its integrity. Security policies draw the fine line between having the most information in play and having too much. Composite services like travel websites must pass data around between multiple corporate entities; they must do it seamlessly, and they must do it safely. The customer's identity must be handed around yet not be subject to confusion, alteration, unauthorized exposure, other mischief.

Passing around information to coordinate multi-party transactions, while necessary, is a source of risk. Sharing financial details necessary to organize a multi-party transaction can easily undo any amount of good practices in-house. But not sharing information creates inefficiency. Consider, for example, a user visiting a travel website. She is probably interacting with several websites, and her personal information is being passed around and shared with each of these. This is efficient and convenient for users but this efficiency and convenience requires trust. Trust is efficient because having to prove trustworthiness in each case is so inefficient.

But trust can be dangerous as well. The problem is finding a balance between the inefficiency of not trusting enough and the danger of trusting too much. In the physical world, we attempt to strike this balance by tolerating a certain degree of transitive trust – "Any friend of Bill's is a friend of mine." Such transitive relationships are usually quite open; that is why they work so well. Transitive trust occurs in the digital world as well, but it is often implicit and generally quite difficult to detect.

Consider the travel website again. If you provide your credit number to the travel site, you may not even question whom the site gives that information to, and even if you do, you will not be able to independently evaluate the trustworthiness of these partners. You trust the travel site to use your credit card number carefully and to only pass that number on to other trustworthy sites. What happens, though, if that trust is broken? Who do we blame and where do we place liability? Placing liability on the deepest pockets won't do in the digital world. When a credit card association like MasterCard allows acquiring banks to obtain credit card financial information in full it exposes itself to security failures at that third party processor – and if that processor fails to protect that data, whose fault is it? MasterCard deals with that uncertainty by imposing its own private body of regulation on its third party processors. Because it cannot easily measure how secure it is, MasterCard tries to eliminate all known faults.

To the extent that trust is built on identity, assertions of identity must be testable across functional and corporate boundaries. Each party to a transaction undertaken in the name of an individual needs to know that individual's name and to be able to test that it is authentic and authorized.[13] This task must be carefully managed. One industry strategy has been to patronize data aggregators, such as credit bureaus. They make good checkers because they know so much more than you so they can give you a trust rating.[14] Aggregating data improves the inferences that can be made from it while at the same time it creates risk that the aggregated data is itself subject to theft or modification.

For data aggregation to be most useful, the infrastructure of its use must itself be shared. Yet if any one member of a shared infrastructure does a particularly good job at risk reduction then that member's residual risk due to shared infrastructure will be dominated by risks caused by the other parties. Is there a moral hazard here? Is there some as yet undesigned underwriting process that can adequately take into account counterparty risk? These are hard questions. In the physical world, contracts supplement norms of human trust; the value of contracts occurs precisely things go wrong. In the digital world, however, is it is not obvious where we should rely on trust mechanisms and where we should rely on enforceable contracts; our intuitions about the physical world are not a reliable guide.

C. Recording and Records

In a free society, it is impossible to enumerate all the things that one is permitted to do. Instead we attempt define what is impermissible. This system creates fuzziness; there is a zone of uncertainty between the obviously allowed and the obviously disallowed. Lest this gray area be exploited, we temper the freedom it offers with a strong notion of accountability. That is, unless the law is so unclear that it offers no guide to the well-intentioned citizen, we leave it to individual citizens to ensure that they are on the right side of the law; if they move too far into the gray we may even place the burden of proof on them to demonstrate that their actions are legal. A good example is the IRS auditing authority. Even if a taxpayer's deductions might ultimately be considered legal the fear of an audit keeps many citizens far within the zone of the permissible.

Because much cyber crime takes place in the gray area between what is clearly permitted and what is clearly forbidden, accountability must play an increasing role in preventing it and ensuring safety in general. What records must we keep to enable meaningful accountability, and to determine which entity is responsible for which problems? After we have determined what kind of records we need, how do we gather them? This is by no means an easy task; digital services and hence digital crimes may be located simultaneously within the facilities of many different owners and, consequently, many different jurisdictions. Adding a third layer of difficulty is the multi-jurisdictional nature of the Internet. Even if we know what records we need and even if we are able to assemble and complete these records, these records are rendered useless if they are not shared across jurisdictional lines. Susan Brenner has looked long and hard at the problem of sharing and accessing records. Her solution is to create inter-jurisdictional search warrants endowed with the full faith and credit of the issuer in the server's jurisdiction.[15] This solution recognizes the degree to which crimes involving computers and networks tend to have none of the locality or physicality of traditional crimes.

A second issue is how long records should be retained. Again, the answer to these questions may be counterintuitive. Paper records take up space, and space is a precious resource in physical world. But not so in the digital world. Available electronic storage space has increased astronomically to the point that neither volume (of data) nor cost (of storage) are inhibiting factors. Simultaneously, in the digital world it is much easier to create records; thus the sheer volume of information has increased dramatically. The combination of these two has made it cheaper to keep everything; the labor involved in sorting through information and making individualized choices about what to save and what to delete soon becomes greater than the cost of keeping everything or nothing.

Keeping everything might not seem like such a bad thing, but corporate firms are understandably hesitant to do so because of the many new liabilities imposed on corporate governance. Consequentially, many firms have an internal policy of destroying all records whose preservation is not mandated by law. Other firms, noting the Morgan Stanley case,[16] conclude the opposite and keep everything so as not to have to defend routine destruction as some kind of conspiratorial activity. In either case, digital physics redirects our policy inquiry. Space is no longer an issue in the digital world, thus the question is not what *can* we keep, but what *should* we keep?

D. The Use of Data Versus its Existence

Zero-cost data acquisition challenges our notions of reasonable expectations of privacy. Much public discussion about privacy focuses on limiting the collection of certain data. In response, the law has a rich history of dictating what data must be, or must not be, collected. [17] But what is reasonable for a person to expect when digital physics is counterintuitive? What is a reasonable expectation of privacy for a general population that does not understand how technologically feasible it is to gather, aggregate and disseminate information? Society does not need laws that forbid impossible actions, hence, there are no rules against data collections that are, or rather were, impossible. But in the digital world the technically "impossible" rapidly becomes technically possible, meaning that the ability to act will always precede any political motivation to regulate that action. Lawmakers are thus inevitably faced with a need to make policy decisions about data collections after such collections are already well established.

Therefore, policy discussions must be centered not on the existence of collectable and observable data but on the *use* of that data. If the price of collection, storage, retrieval, and analysis of data is effectively free, these practices will occur and they cannot effectively be outlawed. Hence legislatures can act only on what people do with the results of the process. For example, it is unlikely that legislatures will ever attempt to criminalize taking a photograph in a public street. Nevertheless, by continuously taking photographs in the public street one could conclude that one's neighbors are never home on Tuesday nights. In like fashion, one could collect the make and license numbers of undercover police cars from information observable by anyone in the public street. Sean Gorman, a Ph.D. student at George Mason University, assembled all available maps on the Internet into one map, leading several people to label his work a serious security threat.[18] The key question for policy makers is not whether data should be collected but what limits we can (constitutionally) place on its publication and use.

E. Network Boundaries Have Zero Alignment with Political Boundaries

As everyone knows, on the Internet network boundaries are not political boundaries (Indeed, for this very reason China has tried to control information flows in and out of the country).[19] Hence it is equally logical to say that the Internet is outside all jurisdictions (hence no ones' laws apply) and to say that the Internet is inside all jurisdictions (hence everyone's laws apply). The former assumption appears in Barlow's "Declaration of the Independence of Cyberspace"[20] and Hughes' "A Cypherpunk's Manifesto."[21] The latter assumption will seem familiar to any global business, especially financial institutions, who regularly find that by having no real location they must conform to the laws of every country.

If we decide that no one's jurisdiction applies, then the Internet will resemble the high seas before 1930. While certain pockets of the Internet remain secure, it is still dangerous to travel between those good neighborhoods when bad neighborhoods lie in between. At the very least, without the equivalent of the Conventions on the Law of the High Seas, there will be no effective or binding control over spam, host-hijacking, or pornography. On the other hand if we assert that the Internet is within all jurisdictions, we invite the problem of conflicting laws. The League Against Racism and Anti-Semitisms suit against Yahoo in France, [22] and the WTO's finding that American prohibitions against Internet gambling violate the rights of Antigua and Barbuda,[23] illustrate the point.

Assuming that we want some jurisdiction's laws to apply, the next question is how to define the boundaries of the relevant jurisdictions. There are two possibilities. Either the boundaries of the physical network—e.g., the ends of the wires, the locations of satellite uplinks, and the physical locations of servers—define the jurisdictions or the jurisdictions define the boundaries. If we choose the former approach, we might pick, for example, all the property of Verizon or all the computers with names in the *.edu domain to be the boundary. If we choose the latter, then we might declare that a citizen of France is bound by French Internet Law regardless of where they live in the world or how it is they get online. Here in the United States, skirmishes over who has the right to tax interstate transactions made on the Internet illustrate how the boundary and the jurisdiction are in considerable tension.

One could imagine a solution to this choice of law problem where every website includes a consent form or a disclaimer that reads, for example, "by choosing to visit this website you agree to be bound by Brazilian law." But this, of course, raises a different question: Why shouldn't a country be allowed to protect its citizens from the actions of foreign websites? Another solution, although one fraught with unintended consequences, might be to create a single, international jurisdiction for the Internet, governed by some international body such as the United Nations. This type of solution is being hotly debated at least in one area: control of the Internet Corporation for Assigned Names and Numbers (ICANN). The Internet's structure assumes that users rely on names to find the underlying network numbers. Thus control over the translation of names to numbers is a point of leverage that would control everything else. Is there an international body ready to handle such responsibility? Or, as I believe, is international political control of the root nameservers a recipe for disaster?

F. So, where to put policy?

As we noted in our discussion of risk and risk management, every location on the Internet is equidistant from every other. This turns our understanding of the physical world, with its separation between good neighborhoods and bad ones, on its head. In real estate location is everything because good neighborhoods are worth paying for. But given the law of equidistance in the digital world, there is no way to fence off the good areas from the bad. The National Academies of Sciences, for example, recently published *Youth, Pornography, and the Internet*, a book that explores the difficult policy choices that we face when trying to protect children from adult material on the Internet. The authors offer two important insights. First, technical change will perpetually outpace our notions of justice and accountability. That is, the technology will allow us to do many things long before we can decide whether or not we *should* do those things. Second, the interests of adults will perpetually trump the needs of children. My own view is that we should deny adults some privileges when they cannot be limited to adults and will prove harmful to children, but the courts, with a few recent exceptions, have often concluded otherwise.[24]

The question for law makers is whether we should rely on the market to produce a solution or whether we should intervene by creating criminal or civil sanctions. If we choose the latter approach we can either impose regulation and sanctions in the network itself (for example, at the level of the ISPs), or at the end-nodes (that is, at the level of users' computers or network routers). The Internet and its many standard protocols are designed to be "end-to-end."[25] This means that the network between two end-points serves as a delivery mechanism and only as a delivery mechanism; this leaves the nature of any communication between end-points to the end-points to negotiate as they see fit. The freedom this structure created is precisely what enabled the Internet to grow and evolve as it has. Had it not been this way, every design change would have had to go through review by governments, ISP's or some other party whose interests might well have been contrary to those who were innovating.

Forcing the network to implement policy is a seductive idea.[26] It is particularly attractive to regulators because the number of entities one has to regulate are the few (backbone providers) and not the many (individual computer owners). The Internet's end-to-end nature has been the key to its success; regulating networks is a bad idea because it will harm innovation. If the state deputizes networks to regulate on-net behavior, one can be assured that the many unfortunate features of telephone monopolies will be visited on the Internet. Attaching devices to phone lines was strictly forbidden until the early 1970s and even then only if one rented bulky and expensive equipment from the phone company. Only after that monopoly control was broken did fax machines become feasible. If one had to register one's fax machines with the phone company one could well imagine that fax adoption would have been delayed a decade. A more current and contentious example concerns whether network providers should be responsible for file sharing across their infrastructures. Clearly (and in line with the Betamax decision[27]), file sharing has substantial non-infringing uses even if adoption rates are driven by parties partial to infringement.

Nevertheless, protecting the end-to-end design principle presumes competent end-nodes or users. That is, it assumes that users want to make and are capable of making effective decisions about self-protection. If the users are unwilling to make such decisions we may need stronger regulation to provide them with motivation they lack. And if they are not capable of making such decisions, then it is not only futile to hold them accountable for their actions, it also unfair.

Thus, before deciding to place policy at the level of end users, we should closely examine our assumptions about the desires and capabilities of those users. Though the growth in Internet users shows signs of slowing, it still occurs at an impressive rate. If the number of persons with access to the Internet doubles every six months worldwide then at any given time only 1/10th of 1% of all users have more than five years experience with the Internet. Even if we assume that every user with substantial experience with the Internet will make informed decisions about self-protection, the sheer volume of inexperienced users should make us hesitant to place a large amount of confidence in the group of users as a whole.

I am no fan of protecting people from themselves, but I recognize that often people democratically demand to be taken care of. Physical world examples include treatment of diseases that are themselves pure products of unhealthy lifestyles, or the mandates for passive restraint systems in automobiles that followed when it became clear that a majority of motorists would not use active restraint systems. It would thus seem likely that people will demand to be taken care of in the digital world.

Unwanted email (spam) is majority of all email and, of that, a majority is relayed through unsuspecting end nodes. Many of these end-nodes demonstrate that they are incompetent, for example, by refusing to scan for hijacking viruses or by clicking on links that are falsified. Calling this behavior incompetence is not a normative judgment; rather it is to say, as a descriptive matter, that these users are ignorant of the danger their own computers create. As noted earlier, it is natural for users to feel that if their computer is doing what they want it to do then everything is fine. The ignorant user, of course, is unaware that the computer may be doing much more than that. Of course, the average user would prefer that the computer not be doing these things, but the problem is that the user does not even know to check for this hidden undesirable behavior.

If users are incompetent, even blamelessly so, it seems useless to regulate them. But if we do not regulate the users, should we then regulate the ISPs? For example, should we require ISPs to read users' e-mail content before deciding whether to forward it? Should we require ISPs to take responsibility for the security problem of their incompetent users by scaning users' computers for virus infections and taking them off the Internet if they are infected? Should we require ISPs to impose traffic limits on end-nodes that would prevent both file-sharing and e-mail relays (used by spammers)? Answering yes to any of these questions, of course, raises the same chilling effects on innovation discussed above, not to mention dangers to civil liberties.

But if regulating end users is useless because they are incompetent, and regulating the ISPs is too chilling to either innovation, civil liberties, or both, then what is left? Perhaps the focus should be on how to create competent users or at least prevent incompetent persons from becoming users. That is, we might require Internet users to demonstrate their competence, to prove they know something before they get to have what would amount to an Internet Drivers' License. Yet, in a borderless world, who would be the issuing authority and, given technical rates of change, how often would a retest be required? If, as I suspect, access to the Internet is becoming a practical necessity for citizens of modern societies, such a license would probably have to be a "shall issue" license. That is, one would not have to prove competence to have the license issued, but some demonstration of malfeasance or incompetence might be grounds for revoking it.

Adam Smith's "invisible guiding hand" will not push the Internet towards order and security because the players are in a classic prisoner's dilemma: the self-interest of each player leads to a result that is bad for the common good. Consider, for example, the problem of record keeping we discussed above. Each corporation would benefit if every corporation shared logs of attempted external attacks, yet the self-interest of each individual corporation is to keep its logs secret because such logs might contain evidence exposing the corporation to liability. Leaving the Internet to its own devices will thus not ensure any sort of order or safety, and given the rapid growth rate of Internet users, we can no longer rely on a conception of the Internet as solely the province of a well-behaved scientific and military elite. Considering that in the laboratory we double the price-performance value of CPUs every eighteen months, of storage media every twelve, and of bandwidth every nine, it is inevitable that our future is one of every greater volumes of data in ever faster motion. Is data security or at least data regulation therefore the thing we should concentrate on? Is it time to say that the real question is not *should* we regulate, but *where* do we regulate?

.

G. Problems Eventually Converge

The wonderful thing about electronic information is that it costs zero to reproduce. This is also the terrible thing. In the physical world, objects have unique identities and take up unique locations in space and time. If an object is located in one place, it can not be simultaneously located in another. Similarly, if I possess your car, you cannot simultaneously possess it. But space and locality are vastly different in the digital world. Digital physics allows the same object to exist simultaneously in a multitude of different locations and, consequently, it allows the same object to be possessed simultaneously by a multitude of users. This principle of digital physics fundamentally changes the nature of business in the digital world.

Another important principle of digital physics is that bits are bits, whether they are an MP3 tune, your mother's date of birth, or a picture of you holding the front page of today's newspaper. Very different types objects (e.g., a song versus a fact versus a photograph) are indistinguishable at the atomic level—they are each just a collection of bits. Thus, unlike in the physical world, we do not need to know the type of object we are dealing with in order to a protect it. To borrow an engineering phrase, the “problem statement” is the same for any piece of digital information: How do we control that information when we cannot be present to protect that information either in space (because a user can be located anywhere on the network) or in time (because a user can access the digital information or re-create it at any time)?

This statement of the problem produces a remarkable convergence: Either we get both digital rights management and privacy, or we get neither. Digital rights management attempts to prevent one copy of an audio recording from becoming many copies even when the originator of the recording has lost all physical access upon commercial release. Privacy management attempts to prevent one copy of personal information from becoming many copies even when the originator of the information has lost all control of it once he or she entrusted it to another. These are the same solutions at the level of the bits even though at the philosophical level it would appear that most of those who hate DRM love privacy and most of those who love DRM do not find privacy a matter of high concern.

How should we negotiate this tension? Should we grant an ownership right to a (US) social security number? Should we encourage digital works of art and science to be rented in a pervasive, always-on networking environment, rather than being owned as copies? That would permit enforceable contracts, but is that what we want? Or, in the least attractive formulation, do we want to build in digital rights management protections into our hardware itself? Congress keeps returning to this approach as if it were the only solution: The V-Chip, the Broadcast Flag, and many other examples come to mind. Nothing can be as harmful to innovation and ultimately self-defeating as such requirements even if, as the example of region coding in DVDs illustrates, some of them do make it to the light of day.

In each of the above examples, we face the principles of digital physics and how those principles affect where we direct public policy. Maybe the only way to implement policy is hardware design; maybe we should simply acknowledge that in a free world there is always risk; and maybe we should stop assuming that every hard problem calls for a technology solution and not a behavioral one. Let us explore this last possibility.

H. Designing for Failure and Creating Accountability

Digital/information goods and services tend to fail suddenly rather than wearing out the way a pair of shoes might. Hence, losses caused by those failures will be sharp and sudden. Once you crack a code-- as De-CSS cracked the protection measures that prevented DVDs from being copied[28]-- the whole protection program is rendered useless.

Perhaps the greatest question before the law with respect to digital goods is who bears the costs for the risk of failure in digital goods and mechanisms and, in consequence, who owns the liability. The digital arena is increasingly essential to everyday life; yet it contains security faults that produce customer losses with neither warning nor any ability by the end user to prevent them. Should liability for cyber-crimes be placed on those who manage digital goods? That is, should we adopt the strict liability rules that courts and legislatures apply to manufacturers of physical consumer goods? Or should the risk of loss (or liability) be placed on the customer? And if we place either or both risks on the consumer, should negligence be the appropriate standard? For example, if the customer is fully up to date with all patches officially released by the vendor, then should this constitute due care and absolve the consumer of either risk of harm or liability?

During the 1990s the commercial world largely caught up with the military world in the use of cryptography. The replacement of the Data Encryption Standard (DES) with the Advanced Encryption Standard (AES) is a case in point; the AES was chosen after an open public competition between commercial and academic contestants from the entire global technical community and was won by a group of software designers from Belgium.[29] The current decade will see the commercial world overtake the military world in the skills of traffic analysis-- the study of who said what to whom and with what frequency. Traffic analysis will quickly become a commercial reality as invisible and as widespread as surveillance cameras. Traffic analysis is extremely powerful, far more so than non-specialists tend to realize. It is the core of what surveillance is about in the digital age. It is no longer necessary to know what a communication *contains*; traffic analysis allows us to figure this out just by observing the *pattern* of communication. Indeed, traffic analysis and not code breaking is the (US) National Security Agency's true specialty. Code breaking is too expensive unless you are reasonably sure that the message in question is worth breaking, and that question—is this message worth breaking– can be answered through traffic analysis.

Traffic analysis poses genuine problems for policy makers. On one hand, traffic analysis could make it much easier to hold Internet users accountable for their behavior. Every router could, in principle, record the traffic going through it, making easier to identify, and thus investigate, suspicious activity. Traffic analyses would be also useful in reconstructing both cyber-crimes and negligent acts that cause damage. On the other hand, traffic analysis presents many dangers. One can misuse traffic data just as much as any other sort of data.[30] As mentioned above, the cost of data acquisition is effectively zero, and the process of data aggregation is automatable and thus also increasingly inexpensive. Thus to safeguard against misuse and to protect privacy rights we must decide whether we want to regulate the data acquisition, the data aggregation, or the use of the data once aggregated. If we try to regulate data acquisition we face the problem of knowing when data acquisition is occurring. Is it better to require specific lifetimes for data and order its subsequent destruction (as the FTC has apparently decided to do[31])? Or should we rather assume that acquisition and aggregation are occurring (because we cannot tell when either or both is happening) and thus require notice? In effect, the law would say (akin to the European Union's Data Privacy Directive) "You can collect, you can aggregate, and you can use but not without my knowledge."

Conclusion

Our brief survey has illustrated a key point: The digital world is not the physical world: Relying on intuitions derived from the physical world to make policy choices will get us into trouble every time. Digital law is and must be counterintuitive. Because our intuition about the digital sphere can so easily be wrong, it is enormously helpful whenever possible to base our policy choices on solid facts.

This is the last time we will have as much hybrid vigor amongst leadership of the security field and it is the last time we will have as clean a slate to work with. Although the law is a valuable resource for security, more important is the diversity of talent and energy we have now, a diversity we must mine while we can.

[1] Moore D, Paxson V, Savage S, Shannon C, Staniford S, and Weaver N, "Inside the Slammer Worm," IEEE Security and Privacy, v1 n4 p33-39, July-August 2003; see http://csdl.computer.org/comp/mags/sp/2003/04/j4033abs.htm

[2] Berinato S, "Why Wasn't the Witty Worm Widely Worrisome?," CSO Magazine, 14 January 2005; see http://www.csoonline.com/alarmed/01142005.html

[3] Shannon C & Moore D, "The Spread of the Witty Worm," Cooperative Association for Internet Data Analysis (CAIDA), 27 April 2004; see http://www.caida.org/analysis/security/witty/

[4] Firebreaks in forests are optimal when they take advantage of natural terrain breaks and can be serviced.

[5] Some major technical universities do no inbound filtering, delegate security management to individual laboratories, and do egress filtering both for the campus at large as well as on a building by building basis. Most cable operators today block outbound email connections except to the message transfer agents they themselves control so as to prevent high-volume, high-speed spread of spam from within their networks to the outside world.

[6] Written up extensively, e.g., the historical record of Doctor John Snow at the UCLA School of Public Health; see http://www.ph.ucla.edu/epi/snow.html

[7] Staniford S, Paxson V, & Weaver N: "How to Own the Internet in Your Spare Time," Proceedings of the 11th USENIX Security Symposium, San Francisco, August, 2002; see http://www.usenix.org/events/sec02/full_papers/staniford/staniford.pdf

[8] Center for Internet Epidemiology, NSF Cybertrust Award, Stefan Savage, Principal Investigator, University of California, San Diego, 24 September 2004; see http://www.jacobsschool.ucsd.edu/news_events/releases/release.sfe?id=293

[9] Presidential Decision Directive #63, "Policy on Critical Infrastructure Protection", 22 May 1998, since superseded by Homeland Security Presidential Directive #7, "Critical Infrastructure Identification, Prioritization, and Protection," 17 December 2003; see http://www.usdoj.gov/criminal/cybercrime/factsh.htm and http://www.whitehouse.gov/news/releases/2003/12/20031217-5.html respectively

[10] A statistician calls this independent "memoryless" arrival a Poisson distribution. The reader encounters these every day; just because the roulette ball has landed on red five times in a row has no bearing on whether it will land on red the next spin. Just because you were not hit by a meteor yesterday does not change your chance of being hit by one today.

[11] For example, animations of the spread of the Witty worm, as found at http://www.caida.org/analysis/security/witty/#Animations

[12] Steve Cooper, CIO of DHS, as grilled by Rep. Adam Putnam, R/Florida, October 10th, 2001.

[13] Some cryptographers believe you can have trust without knowing the identity the person/company you are trusting; for perhaps the best example, see the work of David Chaum and, within that, perhaps his invention of the "blind signature" in 1982 (Chaum D, "Blind signatures for untraceable payments," Advances in Cryptology - Crypto '82, Springer-Verlag, p199-203). For an explanation, see http://www.rsasecurity.com/rsalabs/node.asp?id=2339

[14] This is, of course, a risk management tradeoff inasmuch as the data aggregator can put too much information into play; see the Bob Barr's FindLaw commentary on the ChoicePoint matter at http://writ.news.findlaw.com/commentary/20050415_barr.html

[15] Brenner SW: "Full Faith and Credit for State Search Warrants Subpoenas and Other Court Orders," Working Group on Law & Policy, National Institute of Justice, Electronic Crime Partnership Initiative, 26 August 2002; see http://ecpi-us.org/FullFaithnCredit.html

[16] "Age of Discovery: How Morgan Stanley Botched A Big Case by Fumbling Emails," Wall Street Journal, page 1, May 16, 2005; see http://online.wsj.com/article_print/SB111620910505034309.html

[17] U.S. Supreme Court: OLMSTEAD v. U.S., 277 U.S. 438 (1928), which found no equivalent to trespass in interception of telephone calls hence such interception did not require a search warrant, which was eventually reversed by ALDERMAN v. U.S., 394 U.S. 165, 175 & nn. 8, 9 (1969) which found that the expectation of privacy with respect to telephone calls did exist and hence a warrant was required for interception of them. Now that telephone service is moving away from regulated utilities with defined circuits to the simple "peer to peer" connections of two parties over the Internet, Congress has responded with "CALEA," the Communications Assistance for Law Enforcement Act of 1994, Pub. L. No. 103-414, 108 Stat. 4279, defining the existing statutory obligation of telecommunications carriers to assist law enforcement in executing electronic surveillance pursuant to court order or other lawful authorization and requiring carriers to design or modify their systems to ensure that lawfully-authorized electronic surveillance can be performed. At the same time, cellular telephone service providers are required to

provide geo-location of the user of a cellular telephone to within a hundred yards at all times thus begging the question of whether “Where am I?” is somehow not subject to the expectation of privacy that “What am I saying?” has.

[18] Gorman almost had his dissertation seized. No less than Richard Clarke, cyberterrorism czar for both Clinton and Bush, said that his work should be burned and the Washington Post declined to print it, though they did do a story about not printing it; see http://www.washingtonpost.com/ac2/wp-dyn/A23689-2003Jul7

[19] Zittrain J & Edelman B, "Empirical Analysis of Internet Filtering in China," Harvard Law School, Public Law Working Paper No. 62, IEEE Internet Computing, v70 March/April 2003; see http://cyber.law.harvard.edu/filtering/china/

[20] Barlow JP, "A Declaration of the Independence of Cyberspace;" see http://homes.eff.org/~barlow/Declaration-Final.html

[21] Hughes E, "A Cypherpunk's Manifesto;" see http://www.activism.net/cypherpunk/manifesto.html

[22] San Jose District Court, C00-21275, but in the spirit of this chapter see the discussion in Bratt MK & Kugele NF : "Who's In Charge?", Michigan Bar Journal, v80 n7, July 2001; see http://www.michbar.org/journal/article.cfm?articleID=305&volumeID=20

[23] Difficult to get in full, but if interested in reading the WTO decision begin at this URL http://docsonline.wto.org/ and search for the “Document Symbol” WT/DS285/R

[24] To go further, look at court cases as to whether pornographic web sites available in publicly funded facilities represent a hostile workplace under Title VII of the Civil Rights Act, the various attempts to mandate filtering and the challenges thereto, the record of zero Federal obscenity prosecutions during the whole of the Clinton administration, the many attempts to do something about pornographic spam e-mail, and the Ninth Circuit’s decision that an image on a web page is to be assumed fabricated until proven real – notably that a depiction of a child engaged in sexual acts is not a child unless said child can be provably identified by name.

[25] Saltzer JH, Reed DP, & Clark DD: "End-to-end Arguments in System Design," ACM Transactions on Computer Systems, pp. 277-288, 1984.

[26] Kruse H, Yurcik W, & Lessig L: "The InterNAT: Policy Implications of the Internet Architecture Debate," Telecommunications Policy Research Conference (TPRC), September, 2000; see http://www.tprc.org/abstracts00/internatpap.pdf

[27] Sony v. Universal Studios 464 U.S. 417 (1984), which states that, "...the sale of copying equipment, like the sale of other articles of commerce, does not constitute contributory infringement if the product is widely used for legitimate, unobjectionable purposes. Indeed, it need merely be capable of substantial non-infringing uses."

[28] CSS stands for Content Scrambling System, so De-CSS stands for converting the otherwise scrambled content into viewable content, and that is the name of a somewhat notorious program for breaking DVD protections. The matter began on January 20, 2000, when US District Judge Lewis A. Kaplan of theSouthern District of New York issued a preliminary injunction in Universal City Studios et al. v. Reimerdes et al., in an action under 17 USC 1201(a)(2), also known as section 1201(a)(2) of the Digital Millenium Copyright Act. For a discussion that compares programs to text, see http://www.cs.cmu.edu/~dst/DeCSS/Gallery/

[29] Sponsored by the National Institute for Standards and Technology, the ultimate selection was the Rijndael algorithm from the University of Leujeuven; see http://csrc.nist.gov/CryptoToolkit/aes/rijndael/ and/or http://www.esat.kuleuven.ac.be/~rijmen/rijndael/

[30] A likely apocryphal tale is that the reason certain telephone companies, e.g., the French PTT, do not send an itemized bill is that an itemized bill allows one spouse to know whom the other spouse is calling even if the call itself was unrecorded. (Use your imagination.)

[31] Disposal of Consumer Report Information and Records, Federal Trade Commission, 16 CFR Part 682 effective 1 June 2005; see http://www.ftc.gov/os/2004/11/041118disposalfrn.pdf