DAN GEER

a quant looks at the future

Dan Geer is a security researcher with a quantitative bent. His group at MIT produced Kerberos, and a number of startups later he is still at it—today as Chief Scientist at Verdasys. He writes a lot, and sometimes the output gets read, such as the semi-famous paper on whether a computing monoculture rises to the level of a national security risk. He's an electrical engineer, a statistician, and someone who thinks truth is best achieved by adversarial procedures.

dan@geer.org

THE FUTURE IS ALREADY HERE—IT'S just unevenly distributed [1]. To see some of security's future we do trend analysis on what we already know. This essay demonstrates what can be gotten from open-source intelligence of the most general sort and how it may apply to looking at the near- to medium-term future of security. As with any such work, there are limitations to the method and the results can, if one insists, be pushed too far.

Security is not an end, it is a means. As a technique, it is a form of risk management and a subset of reliability. Real risk management means good decisions, good decision-making requires good decision support, and good decision support requires ordinal scale ($X > Y$) metrics—often no more than ordinal scale.

Where does trend analysis come into this? Trend analysis is what a statistician will recommend when the underlying topic of interest is new and/or changing rapidly and where the method of measuring it is uncertain. In such a circumstance, and so long as the measurement you do have can be applied consistently, the trend data from the measurement can be relied upon even if the raw numbers the measurement returns are suspect. As trends are generally sufficient for decision support ($X > Y$), we've explained why we are here. By analogy, a street cop may never know how much crack is for sale, but he or she can tell a lot from the rise and fall of the street price—enough to make decisions.

Making decisions early is often regarded as something valuable. In the present context, it is good to remember that early decision-making is itself a tradeoff: Making decisions early is more expensive in decision cost than making them later, because early on the choice set is larger and the uncertainty around that choice set is higher. Making decisions later generally comes with fewer workable options, so decision cost per se is less. Trend analysis can thus help you decide not only what decision to make but also when to make it. These are Good Things.

Gather Ye Numbers Where Ye May

There are two sources of numbers: reports from instrumented collection points and surveys. Both may be done by others, and so we must hope that

the ways this data is collected is consistent over time. Surveys are hard to do really right, as they are subject to lots of biases, but the biases are not terribly relevant to trend analysis if those biases are consistent over time. Let's start with the well-known CSI/FBI annual survey [2] and look at the question, "Did your organization experience unauthorized use of computer systems in the last 12 months?" The question has been asked for several years, so there is something to look at (see Figure 1).

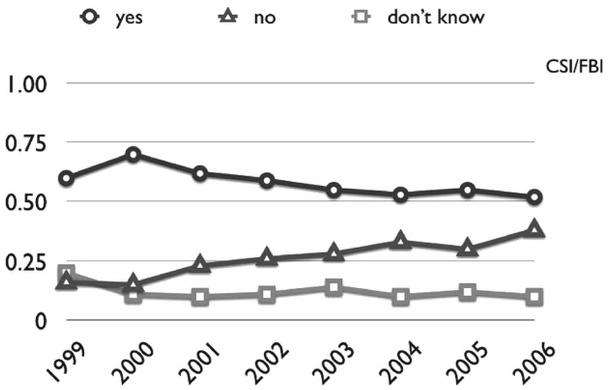

FIGURE 1: UNAUTHORIZED USE IN THE PAST 12 MONTHS

This first trend immediately shows that careful interpretation is part of the effort. In this case, you could say that people who've no idea whether they had or did not have an episode of unauthorized use actually did or did not have such an event. On the one hand, we can say that unless you know you had an event then you did not ("optimistic"), while, on the other hand, we can say that unless you know you did not have an event then you did ("pessimistic"). This is illustrated in Figure 2.

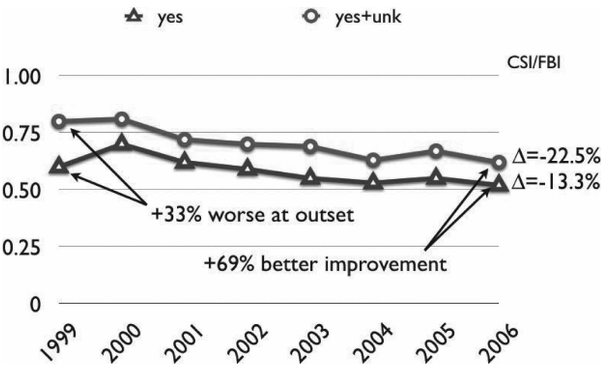

FIGURE 2: OPTIMISTIC V. PESSIMISTIC

This interpretation allows us to think about the problem a little bit deeper, since we now have the upper and lower bounds of assumption, given the data we do have, and we're reminded that the student who gets all As is never the student who gets the "Most Improved" award.

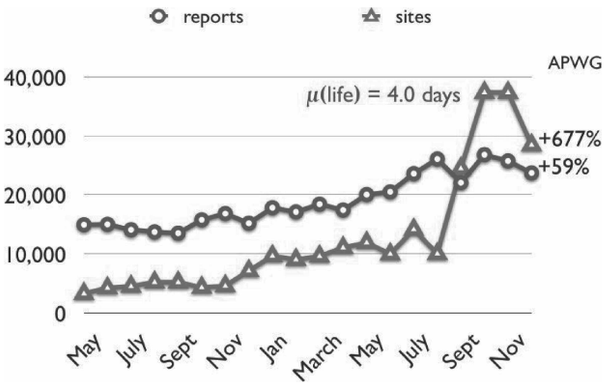

FIGURE 3: NEW PHISHING MESSAGES AND SITES PER UNIT INTERVAL

Let's look at something different: phishing (Figure 3). Data from the Anti-Phishing Working Group [3] shows a 19-month increase of 59% in the monthly reports of phishing email received but a 677% increase in the number of URLs used by phishers in those emails and that the lifetime of those URLs is 4.0 days. This tells us that the supply of URLs is no problem for our opponents and that our opponents cycle the URLs just fast enough to outrun the combined protection bureaucracy response (of consumer to fraud complaint to ISP to hosting center). That tells us that we have lost the supply-side battle, and we should plan accordingly.

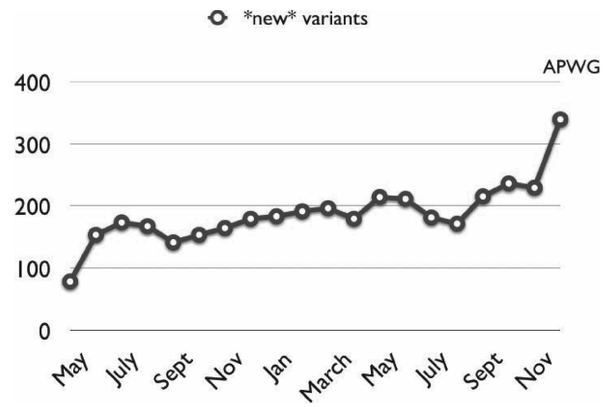

FIGURE 4: NEW DATA THEFT MALWARE VARIANTS IN PHISH EMAILS PER UNIT TIME

These days, phish email often comes with malware attached, and that is certainly a trend worth watching (Figure 4). Although those numbers are not sky-high, it is important as a future-of-security planning exercise to remember that if you are trying to recognize these malware-carrying phish-

ing emails on sight, then your workfactor is the integral of all the phish mails to date, whereas the opposition's workfactor is the price of creating new ones. That, in turn, means that Figure 5 is more like what your force planning exercise has to contend with.

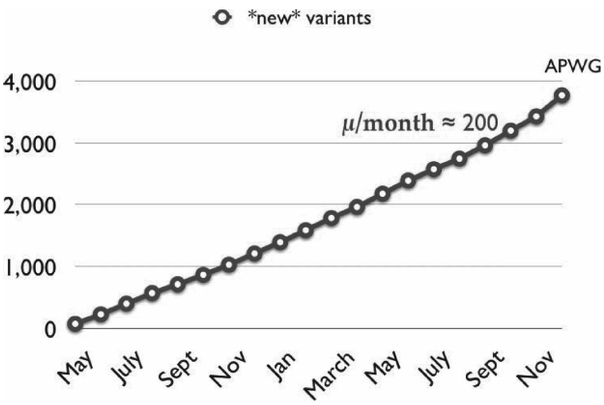

FIGURE 5: CUMULATIVE NEW MALWARE VARIANTS IN PHISH EMAILS

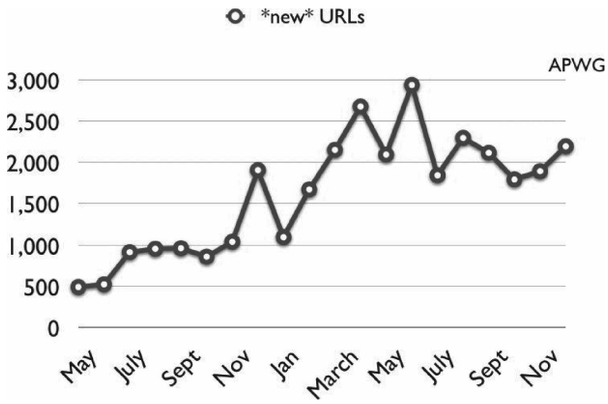

FIGURE 6: NEW DATA THEFT IN URLS IN PHISH EMAILS PER UNIT TIME

Of course, the same thing is true when we look at the URLs that the data theft malware will use if and when that malware succeeds. The month-to-month rate looks like that shown in Figure 6. Figure 7 shows the cumulative effect of Figure 6's rate.

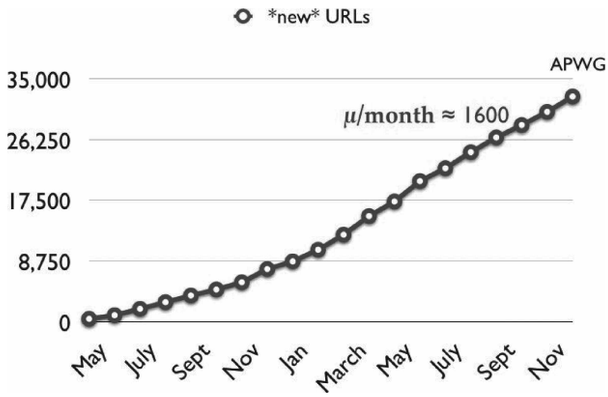

FIGURE 7: CUMULATIVE NEW DATA THEFT IN URLS IN PHISH EMAILS

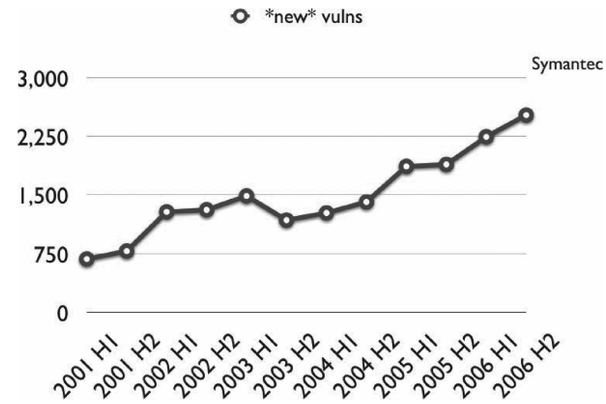

FIGURE 8: NEWLY REPORTED VULNERABILITIES PER UNIT TIME

Let's look at something different—vulnerabilities—and let's switch to Symantec data. Let's also remember that every software vendor is working harder and harder to keep vulnerabilities out of its code. In Figure 8, we can nevertheless see that in the most recent six-month reporting period a new high for identified vulnerabilities was reached. Now Symantec has only been publishing this in its Internet Security Threat Report [4] since 2001, and there have certainly been vulnerabilities around since before that. Even so, if we said that only Symantec hears about vulnerabilities and that there weren't any before 2001, we would have, between then and now, a 26-fold increase since record-keeping began (see Figure 9). Cumulative vulns may not be at the top of anyone's agenda but, in truth, vulns never really go away (they just get rarer, like car owners who never answer recall notices).

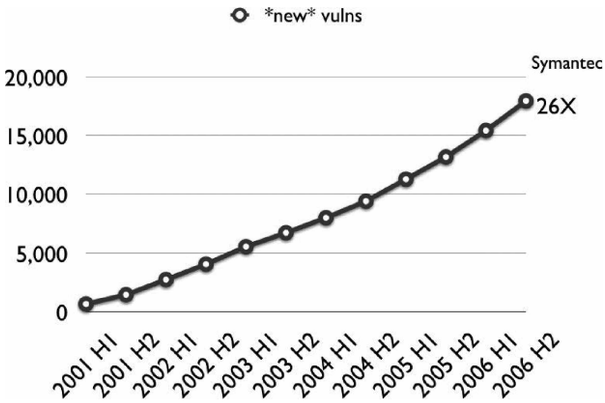

FIGURE 9: CUMULATIVE NEWLY REPORTED VULNERABILITIES

	2005	2004	2003	2002
OS	19	140	163	213
Net Stack	1	6	6	18
Non-Server App.	229	393	384	267
Server App.	88	345	440	771
Hardware	0	20	27	54
Protocol	12	28	22	2
Crypto	0	4	5	0
Other	0	10	16	27

TABLE 1: REMOTE VULNS REPORTED TO/BY NIST

But perhaps you are more interested not in total vulnerabilities but merely in remotely exploitable ones ("remotes"). In that case, NIST has some data for you [5], as summarized in Table 1. Table 1 is exactly as it was reported originally, but as a table it is not as informative as it might be. (Nonserver apps are, by the way, client tools such as Web browsers and email readers.) A better presentation is that of Figure 10 (in which the timeline goes from left to right and mass is displayed as area).

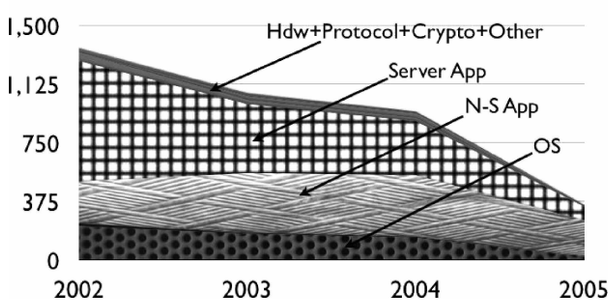

FIGURE 10: REMOTE VULNS BY SOURCE OVER TIME

Hardware	–73.5%
Other	–66.7%
Net Stack	–61.8%
OS	–55.3%
Server App.	–51.5%
Non-Server App.	–5.0%
Protocol	81.7%
Crypto	n.a.
Overall	36.0%

TABLE 2: COMPOUND ANNUAL GROWTH RATE (CAGR) BY REMOTE VULN TYPE

But although the display is more informative, it still isn't good enough. Perhaps it would be better to compute a compound annual growth rate for the various kinds of remote vulns, as listed in Table 2. Now that is more informative, especially as it tells you where progress is being made and where it is not. This might tell you how to redeploy your efforts, for example, but there is yet one more way to look at this, and that is as market share rather than counts. We first construct Table 3 and then use Table 3 to construct Figure 11, where it is now apparent that the action is becoming almost entirely about the nonserver applications. This is important for planning purposes.

	2005	2004	2003	2002
OS	5%	15%	15%	16%
Net Stack	0%	1%	1%	1%
Non-Server App.	66%	42%	36%	20%
Server App.	25%	36%	41%	57%
Hardware	0%	2%	3%	4%
Protocol	3%	3%	2%	0%
Crypto	0%	0%	0%	0%
Other	0%	1%	2%	2%
	100%	100%	100%	100%

TABLE 3: COUNTS OF REMOTE VULNS EXPRESSED AS MARKET SHARE

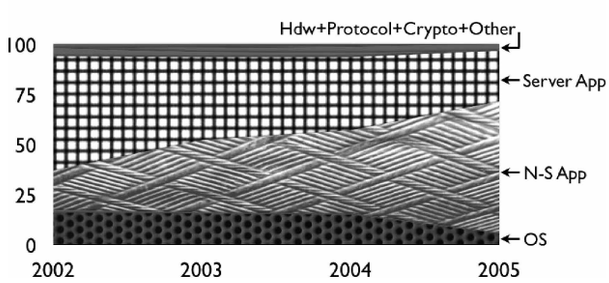

FIGURE 11: REMOTE VULNS BY SOURCE OVER TIME, EXPRESSED AS MARKET SHARE

Let's take a similar look at our very best friend, spam. Because so many people are interested in that topic, we have the luxury of several sources of data. In Figure 12, we have TQM3's take [6] on the volume. In Figure 13, we have Commtouch's take [7] on spam volume and in Figure 14, we similarly have Postini's take [8] on that volume of spam.

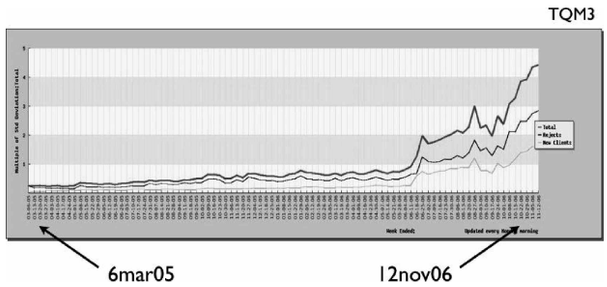

FIGURE 12: ONE ILLUSTRATION OF A SPAM SURGE

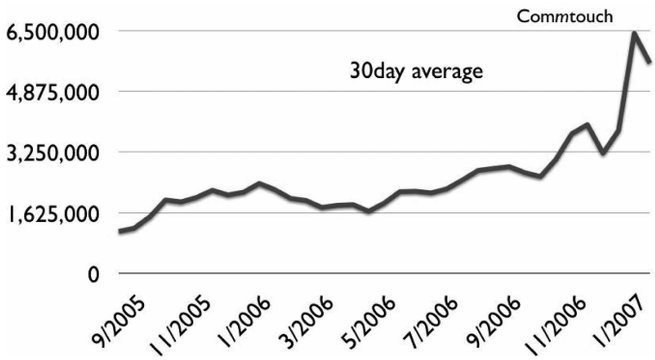

FIGURE 13: ANOTHER ILLUSTRATION OF A SPAM SURGE

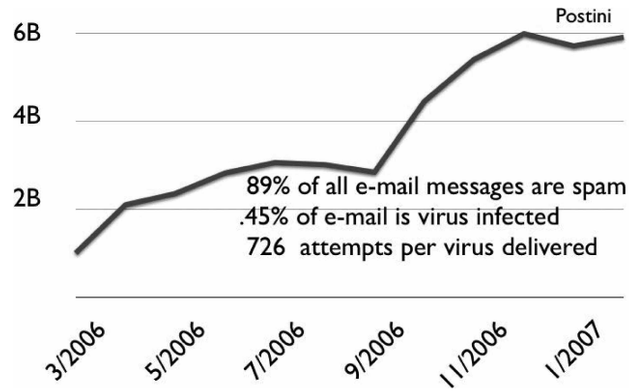

FIGURE 14: ANOTHER ILLUSTRATION OF A SPAM SURGE

It does look as though the trend is upward at not dissimilar rates. Postini's report of additional numbers is itself interesting. For example, it proves that economics lies on the side of the spammer who is trying to get the working attention of the recipient. In the direct mail advertising (junk mail) world, a response rate of 1 in 100 (1%) is considered a success and here we have 1 in 726 for what we can call response rate to virus transmission. The transmitter thus has 1/7 of the direct mail market's definition of success but has that for zero effective cost. The planning information to take from this result is simply that economics favors the opposition, but we also have a metric for how we are doing: whether the 726 number can be made to increase or not.

Incidentally, it is likely that total spam email volume is not rising (despite these three disparate charts) but, rather, that the percentage delivered is rising as template spam (for making individual messages unique) is progressively defeating Bayesian filters.

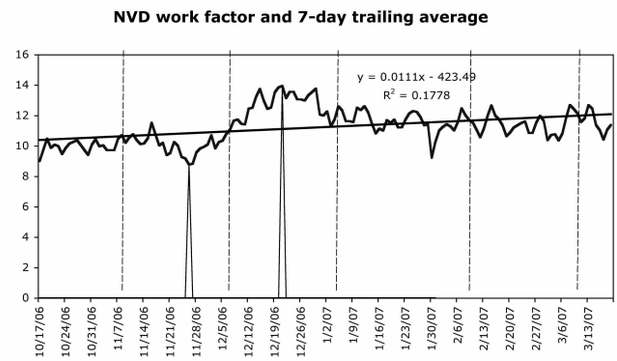

FIGURE 15: WORKFACTOR FOR SECURITY PRACTITIONERS

This is all obviously pointing at work for you, the reader, to do, but how much work? Interestingly,

the National Vulnerability Database folks calculate a daily number—the "workfactor" number. In Figure 15, we see several months' worth as the raw number and a fitted line. Among other things, the fitted trend line is rising, which, as a planning mechanism, says that the workfactor of people dealing with security problems is climbing. The vertical dashed lines are the days on which Microsoft releases its monthly bolus of problems to attend to. The spikes point to the minimum (the Sunday after Thanksgiving) and the maximum (the next-to-the-last shopping day before Christmas). One can be perhaps forgiven for suggesting that this would be consistent with an all-out assault on the Internet Christmas shopper this past season.

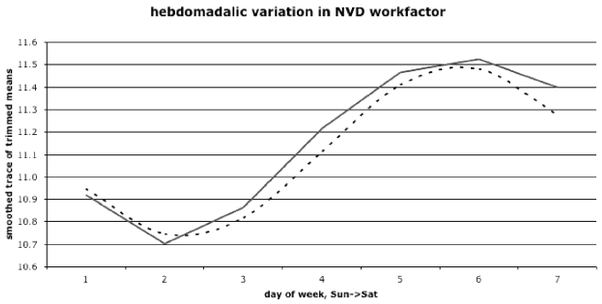

FIGURE 16: EVIDENCE OF A WORKWEEK HIDING IN THE WORKFACTOR DATA

But there is something interesting hiding in these numbers if you look at them a different way: It appears (in Figure 16) that there may be evidence of a conventional workweek. And if the opponent is actually enjoying a conventional workweek, there is perhaps no further need for corroboration that exploiting security problems has become the day job for some number of people. Yes, the dotted line is a sine curve and it does fit pretty well. In fact, we see corroboration of a workweek in Symantec's numbers for the daily appearance rate of unique phishing emails (Figure 17).

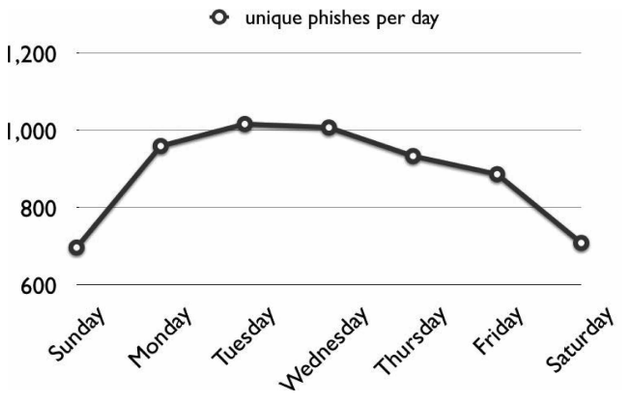

FIGURE 17: NUMBER OF UNIQUE PHISHING EMAILS ON WEEKLY CYCLE

Everyone rightly worries about spyware, trojan horse programs, and especially such nasties as keyloggers. With help from Webroot [9] we can quickly see that if an enterprise PC has any spyware then it probably has more than one example (Figure 18). We can see that trojans are plentiful (Figure 19) and we can see that if an enterprise PC has any trojans then it probably has more than one example (Figure 20) or, even more worrying, that if an enterprise PC has a keylogger then it could well have more than one example (Figure 21).

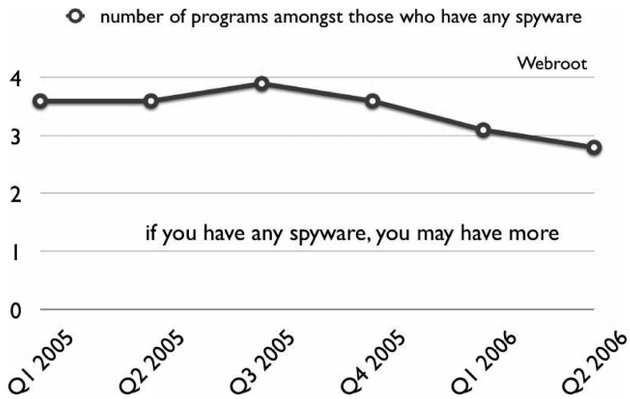

FIGURE 18: NUMBER OF SPYWARE EXAMPLES PER ENTERPRISE PC THAT HAVE ANY

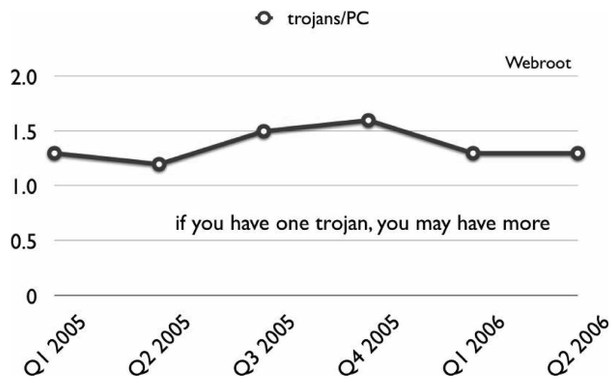

FIGURE 19: PERCENTAGE OF ENTERPRISE PCS WITH A TROJAN

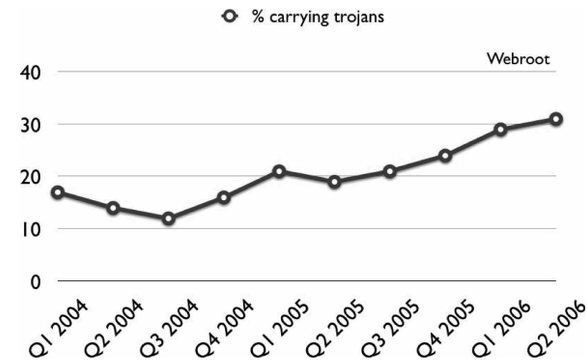

FIGURE 20: NUMBER OF TROJAN EXAMPLES PER ENTERPRISE PC THAT HAVE ANY

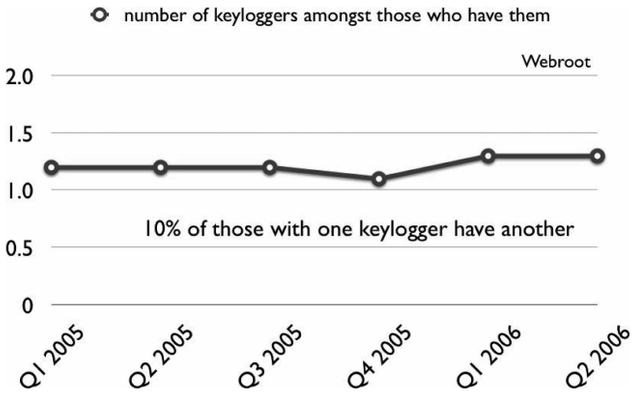

FIGURE 21: NUMBER OF KEYLOGGERS PER ENTERPRISE PC THAT HAVE ANY

This apparent fact makes sense; users who do things that get them in trouble once will probably get themselves in trouble more than once, leading one to concur with Microsoft that having the ability to very quickly re-image a desktop may be an important part of any risk management plan:

> When you are dealing with rootkits and some advanced spyware programs, the only solution is to rebuild from scratch. In some cases, there really is no way to recover without nuking the systems from orbit.
>
> —*Mike Danseglio, Program Manager, Security Solutions Group, Microsoft, April 3, 2006 [10]*

Sometimes, though, you can make better decisions by understanding whether you are a target, *per se*. Using Counterpane's data [11], it is easy to see that where the money is is where the attacks go (Figure 22).

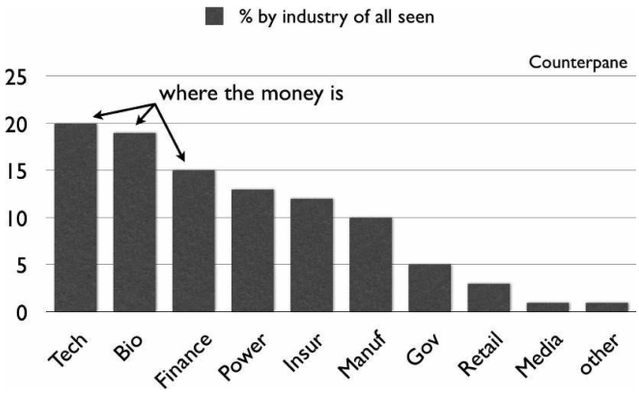

FIGURE 22: WHERE THE ATTACKS ARE IS WHERE THE MONEY IS, AND VICE VERSA

Perhaps your training leads you to think of the great mass of IT in the modern enterprise in the way a public health doctor views infection in a sprawling city. If so, figures like this might make you think:

- 318 new Win32 viruses/week
- 9,163 hosts/day join botnets
- 75% of malware is modular
- 1% of bots show themselves per day
- 5,900 phishing emails/minute

This, too, is part of thinking about the future so as to plan for it. In fact, by this point in this essay, perhaps we can hazard some inferences and identify some implications.

Where This Leads

Security and threat co-evolve, exactly in the same sense that predators are the reason prey diversify. Over time, and as natural immune systems get better, the pathogens that remain are fewer in number but are selected for virulence (the ability to move from host to host), and indeed we've seen that in ever-faster-spreading but ever-rarer epidemics of computer viruses and worms. We've seen that infectious agents rarely cross species boundaries, just like in nature. We know that corruption of the immune system is the worst (think of the "Witty" worm), and we know that parasites co-exist nonlethally with their hosts. (Some wags claim that home machines involved in botnets, except for being 0wned, are better managed than your average home machine.)

We also know that evolution's course is by punctuated equilibria [12] rather than through smooth gradual change. We are living just after such a puncturing of the equilibrium. Public access to the Internet began in 1990, and that access, followed in 1991 by the precursor of the browser, created an irresistible economic force for everyone to connect to this Internet thing. However, doing so suddenly created a world where both prey and predator were and are location-independent. In this new world, force multiplication is proportional to bandwidth, and bandwidth is cheap, almost (but not entirely) too cheap to steal [13]. That opening of the Internet also created the economic driver for commoditization of computers and that commoditization, absent any effective regulatory framework, led inexorably to monoculture and monoculture threat.

For better than three decades, the computing you can buy for a dollar has grown by 1% per week, the storage you can buy for a dollar has grown faster than that, and the transmission capacity you can buy for a dollar has grown even faster still. Over that same interval, total market capitalization, as measured by the Dow Jones Industrial Average, has grown by 1/7% per week, thereby

proving that data comprises a rising fraction of total corporate wealth. Of course, the value thus expressed is a magnitude and, to a large extent, the sign bit is separately determined, in part by security technology and security practitioners. Data has thus become the coin of the realm, being the repository of value for the general economy.

That data is increasingly mobile. The economically optimal computer is changing as we speak. When CPU price/performance doubles every 18 months, storage price/performance every 12, and bandwidth every 9, then for every decade one expects two orders of magnitude in computer power but three orders of magnitude in retained data and four orders of magnitude in data transmission. The implication of spending constant dollars on these three components of a computing infrastructure would thus mean that at the end of a decade the CPU would be only 1/10 as powerful per unit volume of data but the data, despite being 10 times as voluminous, would be able to completely move 10 times as fast. Coupling this with the close embrace of the Internet by commerce at all levels makes it clear that the winners will be those with as much information as possible in play, while the losers will be those who have too much, with security technology and practice providing the fine line between as "much as possible" and "not too much." This is already semi-present, as Gibson would say, with convergence of pure comms (telephony) and data-rich applications.

Data becomes our focus going forward. Security is what distinguishes data that has value from data that does not. Regardless of setting or metaphor, a rising threat requires any defensive perimeter to contract. This is true for the military, for wildebeeste, and for data. A contracted perimeter for data means a shift of focus of our arrayed protection technologies to individual data objects at their point of use. Operationally, data is at risk when it changes from at rest to in motion, a state change akin to evaporation. The point of use is where that state change occurs, and thus monitoring is the first priority because in the electronic world that which escapes your view is that which will escape your grasp (i.e., you cannot control what you cannot see). The single smartest thing any Cabinet Secretary has said in thirty years was Secretary of Defense Donald Rumsfeld's comment that it is the unknown unknowns that will kill you (and every journalist and pundit who made fun of it thus proved beyond doubt that they are innumerate). Security metrics therefore begin with certainty at the point of use.

There are lots of interesting but decidedly losing propositions for how to handle a future that is about data security:

- Perform content inspection. This can be defeated by Pig Latin, much less encryption.
- Use statistical anomaly detection. This defeats itself, as it creates an infeasible work-factor to damp out false positives.
- Look for signatures. Like antivirus programs, this is defeated by any enemy, as it is the Red Queen's own technology, "Around here, it takes all the running you can do to keep in the same place" [14].

No, the trends and the facts tell us that the engineering problem statement now facing us is data protection that is (1) inescapable, (2) invisible, and (3) future-proof. The rules of economics tell us that this is a minimax problem, meaning an optimization tradeoff between preventing trouble (anticipation costs) and cleaning up trouble (failure costs). The National Center for Manufacturing Studies perhaps illustrates this best [15]. Figure 23 shows that near-infinite spending on prevention does get near-zero spending on failure recovery, just as near-zero spending on preventing trouble risks near-infinite spending on failure recovery. The economically optimal point is the sum of the two curves, the minimum cost for the maximum protection—a "minimax" solution. Though not shown, as the degree of electronic collaboration rises, the failure costs at a given level of information assurance will rise, thus pushing the summed cost curve upward and rightward as the essentialness of electronic collaboration grows.

NCMS

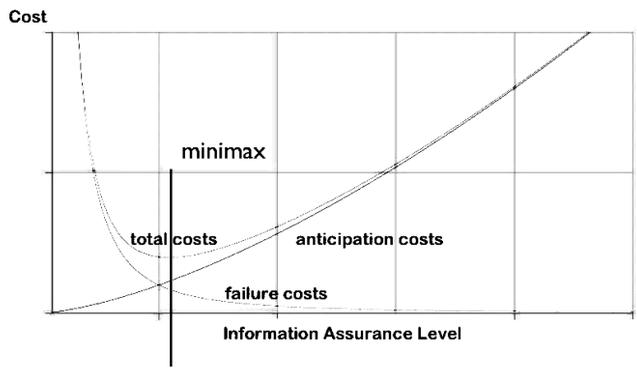

FIGURE 23: BEAR VS. AVOID THEM

Summary

In a sense, this essay should be unsurprising and it should feel unfinished. It is unsurprising, and it is unfinished. The trend data tells us that our opposition is gaining ground in an asymmetric war, a war where our costs accumulate and theirs do not. It tells us that our substantially increased levels of effort in protective armamentarium, in better prevention of vulnerabilities, and in improved detection of all sorts are proving not to be enough, as despite the rise in protective input there is a faster rise still in the capabilities of that which must be protected against. This calls out for the only advice there is: *If you are losing a game you cannot afford to lose, change the rules.* The rules we have to change are what it is we think we are protecting. A lost laptop is economically meaningless besides the data it contains. A single point of failure that must exist for absolute design reasons needs layers of defense-in-depth. Cascade failure cannot be cost-effectively prevented except by diversification when the efficacy of protections is, as these graphs show, falling despite the best efforts of good and honest people. Because data is where the value is, that is where the protections must go. If we are lucky, the worst tradeoffs we get are "DRM and privacy: both or neither." If we are unlucky, we get neither freedom nor security and neither privacy nor convenience, but the unluckiness will be because we failed to make necessity be the mother of invention. The trends are not good, but they are not yet a disaster. All of them have a consistent direction and tilt; what will be a disaster is if that direction and tilt continue, and that disaster will arrive far sooner than global warming.

REFERENCES

[1] William Gibson, author of *Neuromancer*, NPR interview, 30 November 1999.

[2] http://www.gocsi.com/forms/fbi/csi_fbi_survey.jhtml.

[3] http://antiphishing.org/reports/apwg_report_december_2006.pdf.

[4] http://eval.symantec.com/mktginfo/enterprise/white_papers/ent-whitepaper_internet_security_threat_report_xi_03_2007.en-us.pdf.

[5] http://icat.nist.gov/icat.cfm?function=statistics.

[6] http://tqmcube.com/tide.php.

[7] http://www.commtouch.com/Site/ResearchLab/statistics.asp.

[8] http://www.postini.com/stats/.

[9] http://www.webroot.com/pdf/2006-q2-sos-US.pdf.

[10] Mike Danseglio, Program Manager, Security Solutions Group, Microsoft, April 3, 2006; http://www.eweek.com/article2/0,1895,1945808,00.asp.

[11] http://www.counterpane.com/cgi-bin/attack-trends4.cgi.

[12] N. Eldredge and S.J. Gould, "Punctuated Equilibria: An Alternative Tophyletic Gradualism," in *Models in Paleobiology*, edited by T.J.M. Schopf (Freeman Cooper, 1972).

[13] It is actually better to steal that bandwidth, since if you register a block of static addresses, then people will blacklist that block.

[14] L. Carroll, *Through the Looking Glass*, Chapter 2, 1872, replicated in the "Red Queen Hypothesis" in the study of co-evolution of parasites and hosts; for that see L. Van Valen, "A New Evolutionary Law," *Evolutionary Theory* (1973), vol. 1, pp.1–30.

[15] http://trust.ncms.org/pdf/CostInfoAssur-NCMS.pdf.