

A Quant Looks at  
the Future  
Extrapolation via  
Trend Analysis  
Dan Geer 21iii07

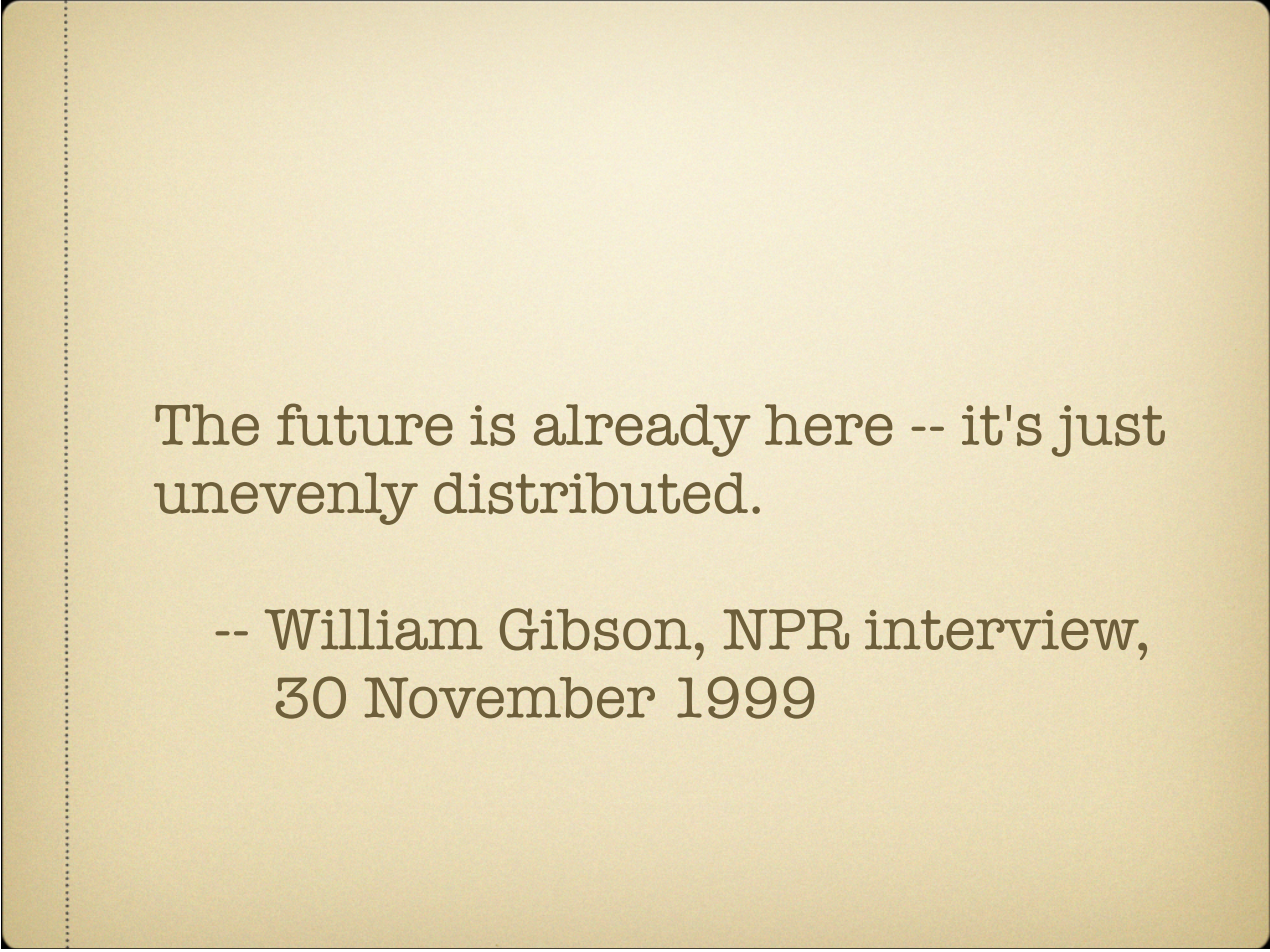
Trends in security, with possible implications.

Feel free to contact for more (there's lots): Dan Geer,  
dan@geer.org, +1.617.492.6814

# Why trend analysis?

- Self-correcting errors (under general conditions)
- The first step in thinking about the future
  - Earlier detection  $\Rightarrow$  earlier control
  - Trades decision cost against narrowed options

Why do trend analysis? First, trend analysis is what a statistician will recommend when the underlying topic of interest is changing and 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 can be relied on and it is what you need for decision support. Of course, making decisions early is more expensive in decision cost than making them later, but then again later decision making generally comes with fewer workable options.



The future is already here -- it's just  
unevenly distributed.

-- William Gibson, NPR interview,  
30 November 1999

That is what trend analysis is all about; futures. Quotation  
appeared in print in "Peering round the corner," The  
Economist, 11 October 2001.

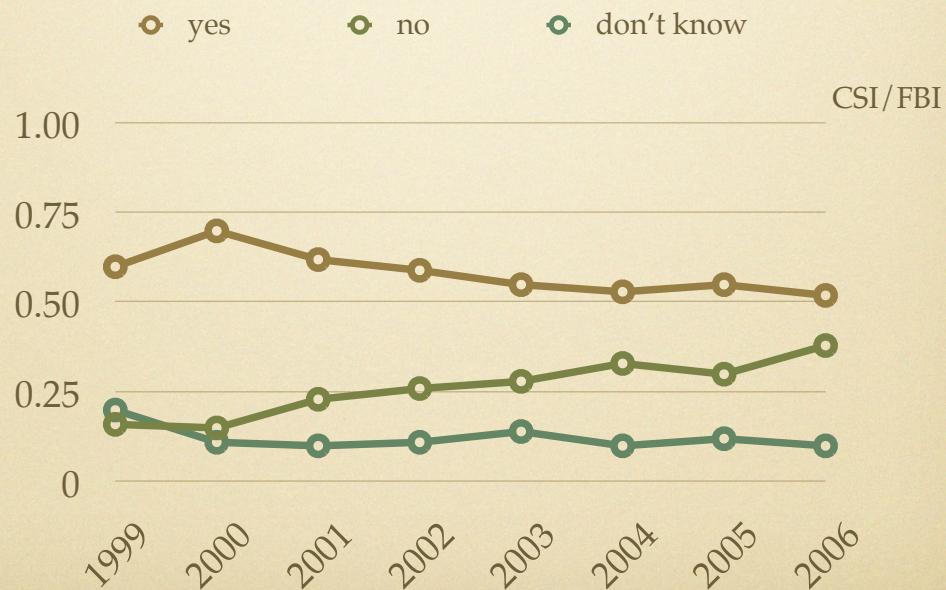
# What is available?

- Primarily two sources / types
  - Event reports at collection points
  - Surveys

And what sorts of security trend data is there in the public domain? Mostly that which comes from collection points for incidents and that which comes from surveys.

Unauthorized use

# Unauthorized use

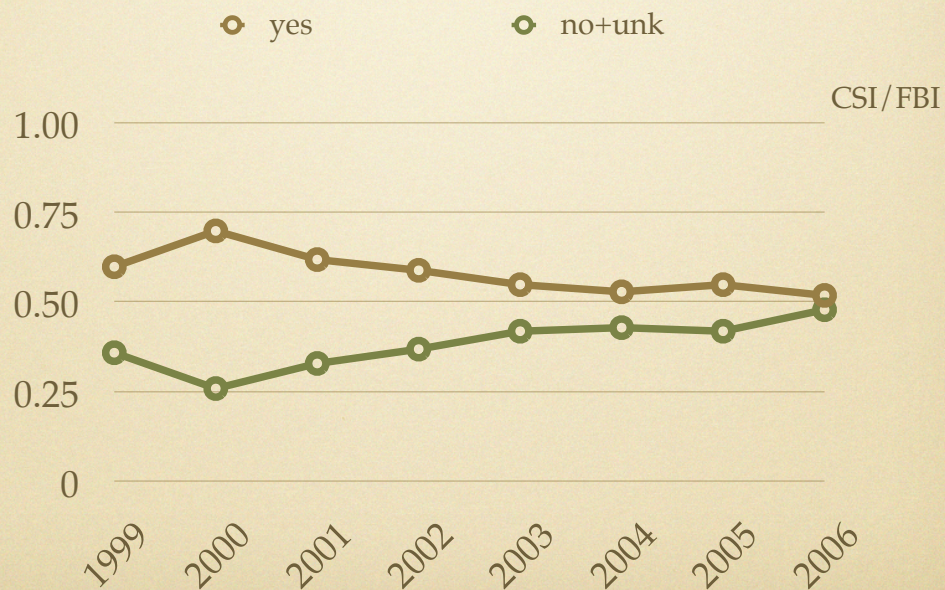


This is the Computer Security Institute and Federal Bureau of Investigation annual report where, for this slide, the question is “Did your firm experience unauthorized use if its computer systems?” and, as you can see here, a majority answered “Yes.” (Reports are released in July for the previous year; see <http://www.gocsi.com/press/20060712.jhtml>)

data

0.6,0.7,0.62,0.59,0.55,0.53,0.55,0.52  
0.16,0.15,0.23,0.26,0.28,0.33,0.3,0.38  
0.2,0.11,0.1,0.11,0.14,0.1,0.12,0.1

# Optimistic



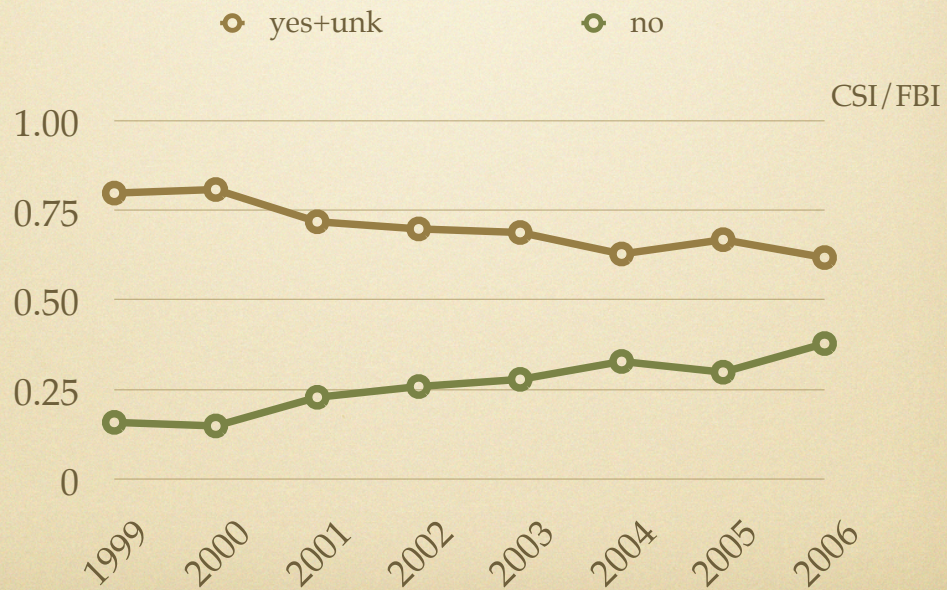
Of course, a bunch answered “I don’t know” so we have a choice. We can be optimistic and assume that people who did not know whether they had been attacked had, in fact, not been attacked.

data

0.6,0.7,0.62,0.59,0.55,0.53,0.55,0.52

0.36,0.26,0.33,0.37,0.42,0.43,0.42,0.48

# Pessimistic



Of course, we can be pessimistic and assume that people who did not know whether they had been attacked had, in fact, been attacked.

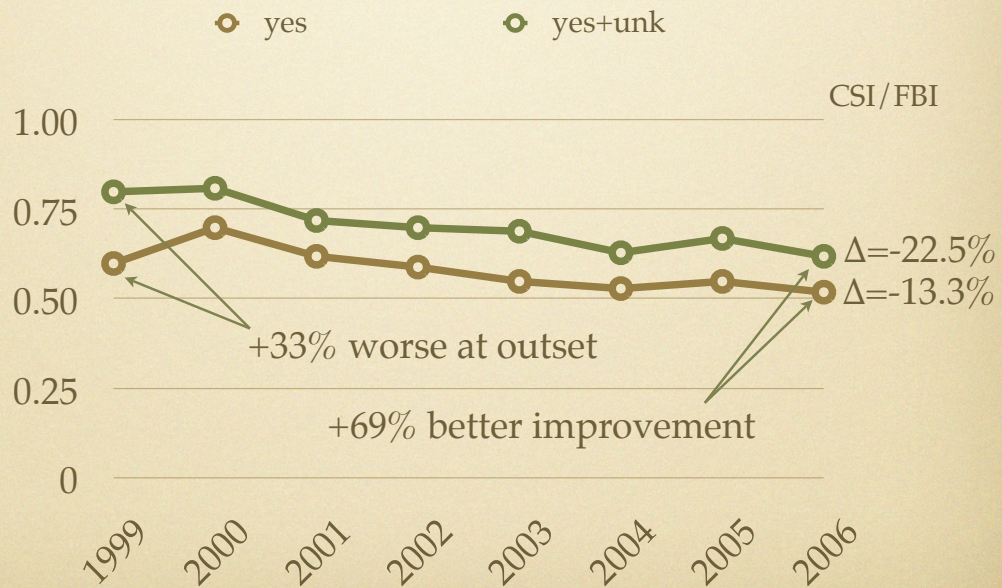
data

0.8,0.81,0.72,0.7,0.69,0.63,0.67,0.62

0.16,0.15,0.23,0.26,0.28,0.33,0.3,0.38



# What'll it be?



Looking at this as trends, then, we have a choice between the optimistic (“yes-only”) and the pessimistic (“yes” and “don’t know”): On the one hand, the pessimistic view gives you a 33% worse picture at the outset but a 69% better improvement over time. And vice versa. This illustrates that trend analysis is both possible with scant data and asks some questions of its own.

data

0.6,0.7,0.62,0.59,0.55,0.53,0.55,0.52

0.8,0.81,0.72,0.7,0.69,0.63,0.67,0.62

- Net: amongst CSI/FBI respondents unauthorized use is falling very slowly
- Caveat: all self-selecting surveys are suspect

While the rate of unauthorized use does appear to be falling slightly, this does require that there be no change in the willingness of respondents to disclose unauthorized use. In that sense, this is a “self selecting” survey (respondents either select themselves as survey participants or they select the manner in which they answer) and self-selecting surveys are problematic statistically speaking. Don’t throw the data out, but be skeptical.

# Incidents

# Incidents, % of firms

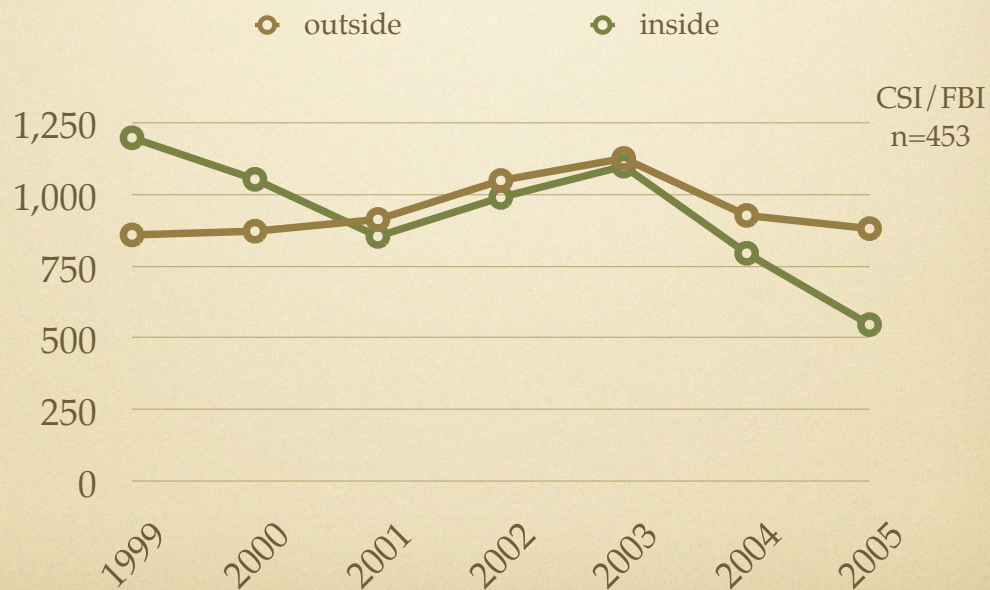


What's interesting about this is that it says insider and outsider attacks are equivalent. (Calculated as 1-last\_col, Table I, p.13, of FBI2005.pdf, i.e., it is the percentage of those who are certain they had an outsider attack and the percentage of those who are certain they had an insider attack.) Unfortunately, CSI/FBI changed how this was asked and thus they no longer report this number in a way that allows this continuing this trend line.

data

0.60,0.58,0.62,0.72,0.69,0.70,0.65  
0.65,0.63,0.59,0.64,0.68,0.66,0.56

# Incidents, est. counts



But it also says that insider attacks are falling as compared to outsider attacks. (Calculated as column percentage times column minimum (1,6,11) summed over the three columns times n(respondents)=453.) Unfortunately, CSI/FBI changed how this was asked and thus they no longer report this number in a way that permits this trend line to be carried forward.

data

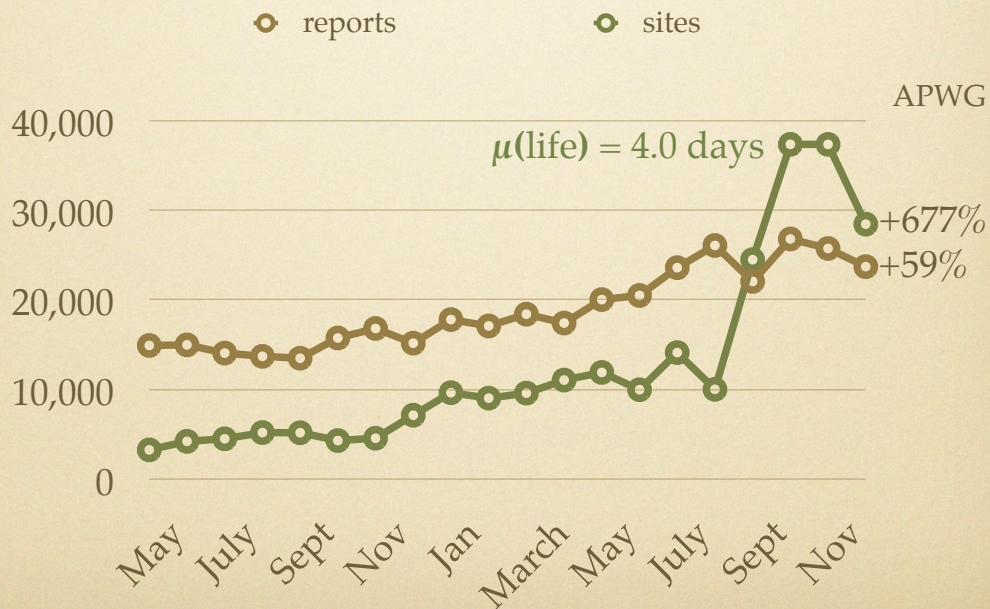
861,874,915,1051,1128,929,883  
1200,1055,856,992,1101,797,548

- Net: amongst CSI/FBI respondents attack rates for insiders and for outsiders are falling but faster for insiders
- Caveat: “[T]he number of respondents willing to report their losses this year was less than half the number of the previous year,” which is why surveys without followup for non-response are problematic.

Same comments as with the other round-up of CSI/FBI numbers: Self-selection requires a careful eye in the face of such comments as in `FBI2006.pdf`, viz., “[T]he number of respondents willing to report their losses this year was less than half the number of the previous year.”

Phishing for data

# Phishing, \*new\* only



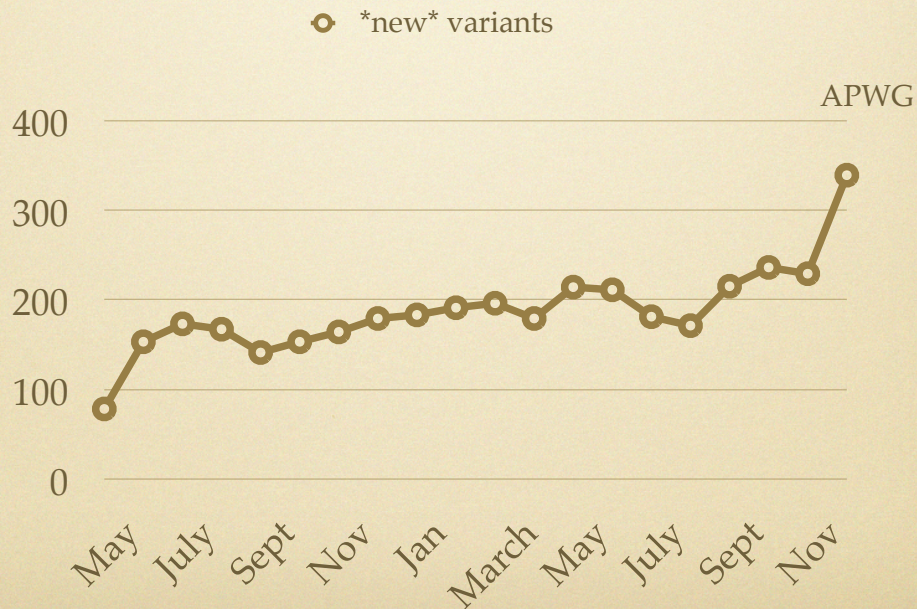
The Anti-Phishing Working Group has some interesting data; what we see here is a 19 month increase of 59% in the reports of phishing e-mail received but a 677% increase in the number of URLs being used by phishers. See [http://antiphishing.org/reports/apwg\\_report\\_december\\_2006.pdf](http://antiphishing.org/reports/apwg_report_december_2006.pdf)

data

14987,15050,14135,13776,13562,15820,16882,15244,17877,17163,18480,17490,20109,20571,23670,26150,22136,26877,25816,23787  
 3326,4280,4564,5259,5242,4367,4630,7197,9715,9103,9666,11121,11976,10047,14191,10091,24565,37444,37439,28531



# Data theft malware

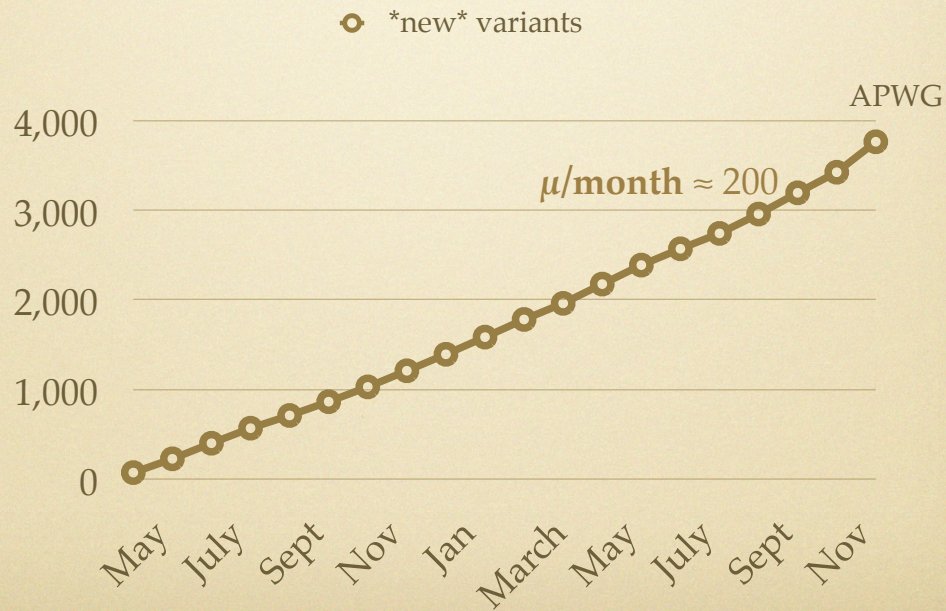


A disturbing part of phishing is that in addition to the automated social engineering that phishing represents, malware is beginning to come along for the ride. We have here the number of new malware variants per month in phish e-mails that APWG has seen. New per month: is there any doubt that variant count will defeat signature analysis?

data

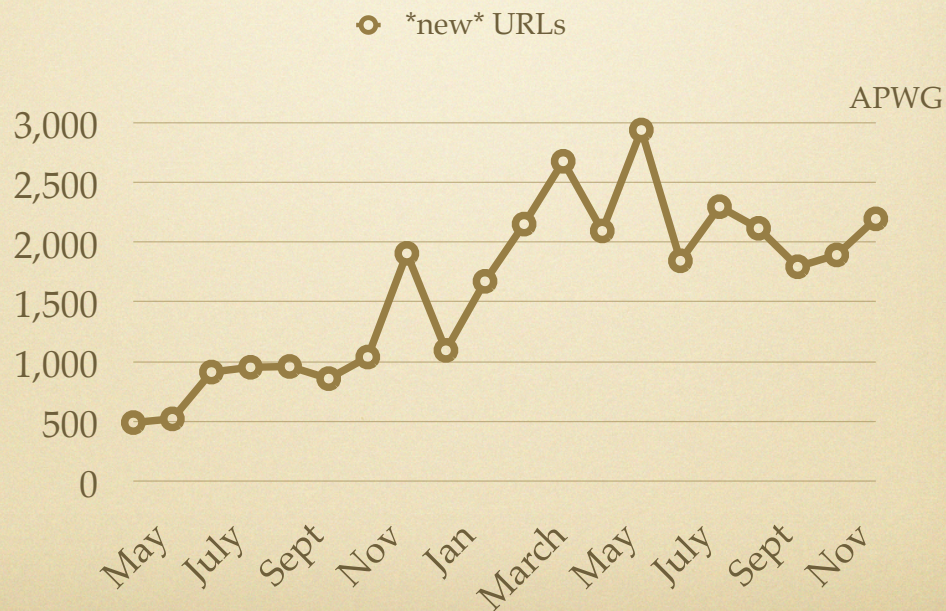
79,154,174,168,142,154,165,180,184,192,197,180,215,212,182,172,216,237,230,340

# Cumulative



So, taking the trend analysis further, we have here the “cumulative” increase, i.e., the number of new malware variants hidden in phish e-mail that have accumulated over the course of 19 months. Pretty steep growth; average per month is 189 (“~200”).

# Data theft malware

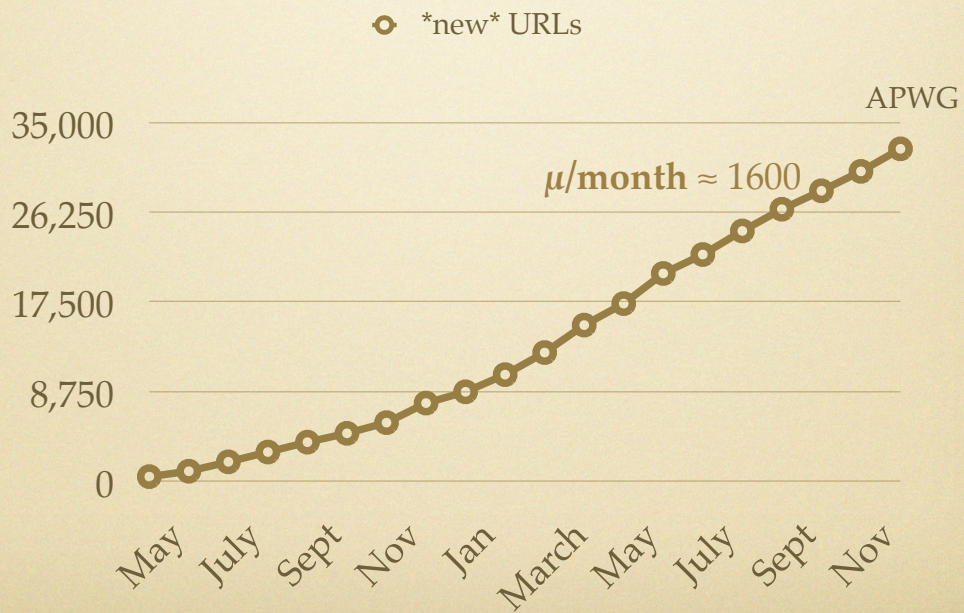


This is the month-by-month number of new phish related URLs seen. Not shown is that such an email lasts on average five (5) days and the longest yet seen was thirty-one (31) days.

data

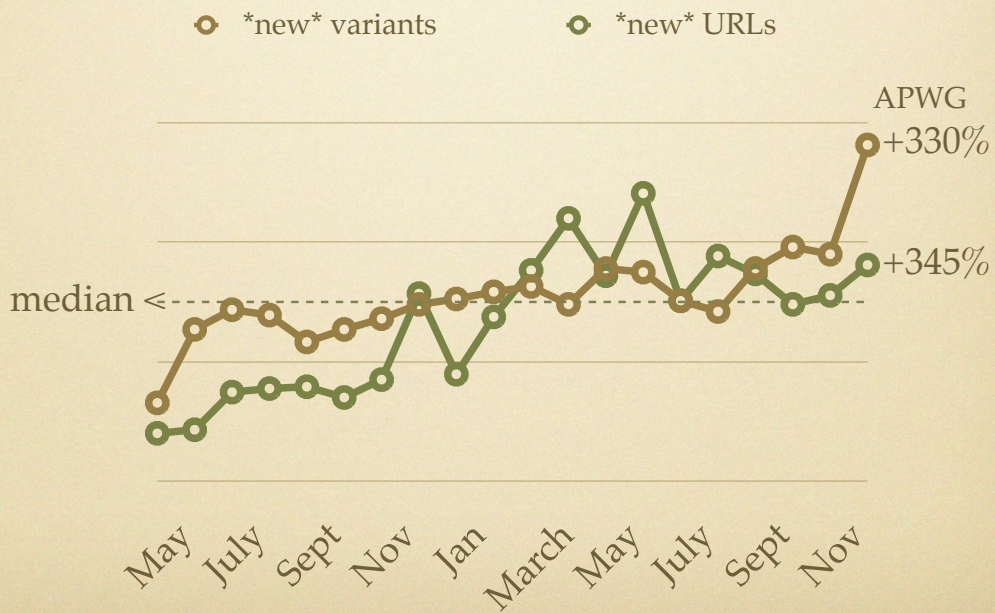
495,526,918,958,965,863,1044,1912,1100,1678,2157,2683,2100,2945,1850,2303,2122,1800,1899,2201

# Cumulative



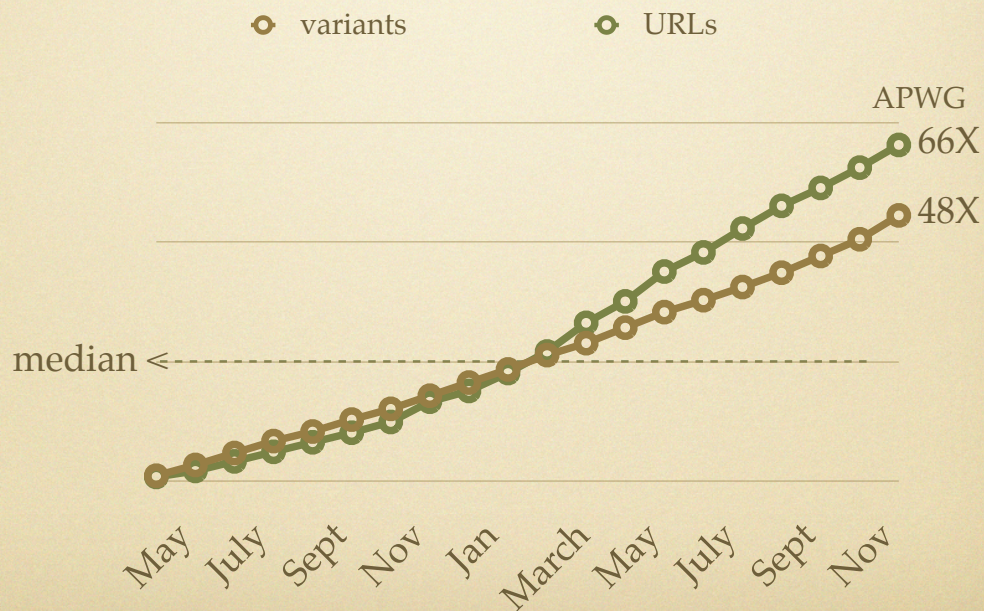
As before, this is a cumulative curve of all the new URLs seen over the course of 19 months; average per month is 1626 (“~1600”).

# Re-scaled to compare



If we rescale (“normalize”) the two curves so that we can superimpose them on each other, then this is what we see. Taking May of last year as if we had never had a problem before, we find a 330% increase in malware variants and a 345% increase in URLs over the past 19 months.

# Rescaled, cumulative



Finally, this is the superimposition of the cumulative curves showing an increase in the number of malware variants (had we started with nothing but the ones found in May of 2005) by a factor of forty-eight (48) and an increase in the number of URLs by a factor of sixty-six (66). The thing to remember is that the defender's work factor is proportional to this cumulative curve while the attacker's work factor is the cost of a new variant. As the latter is now automated, the arms race between attacker and defender can be manipulated by the attacker to bankrupt the defender.

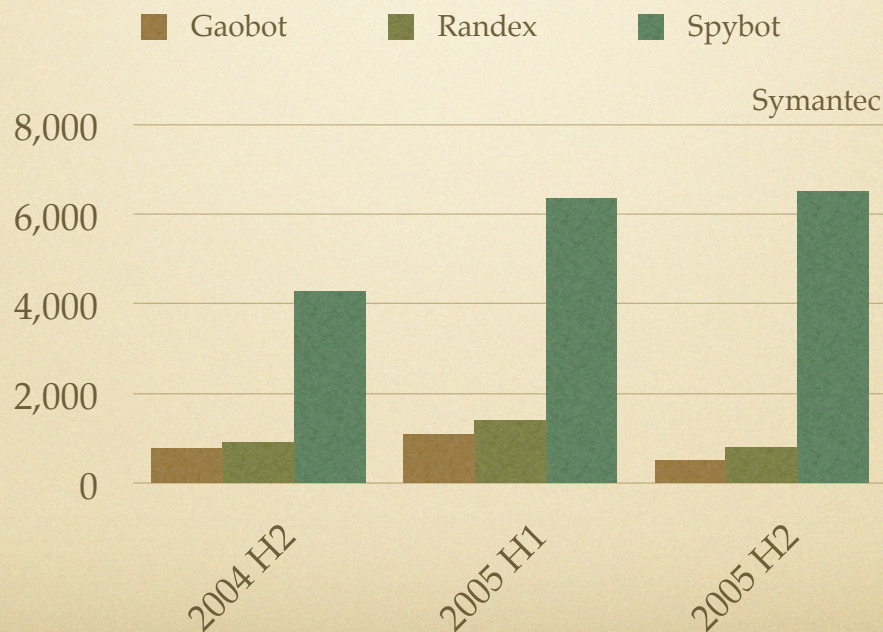
- Net: phishing is professional and it is after data; whether or not the insider / owner is a thug or not is irrelevant when said insider / owner can be made to act like a thug
- Caveat: phishers abandon variants & URLs quickly so the number in circulation is not that cumulative number -- but folks blocking by URL or variant have a work factor like that cumulative number

Phishing is now a professional sport and it is after data. Whether the insider is a thug or the outsider can make the insider act like a thug is totally irrelevant at any level of truth.

Malware

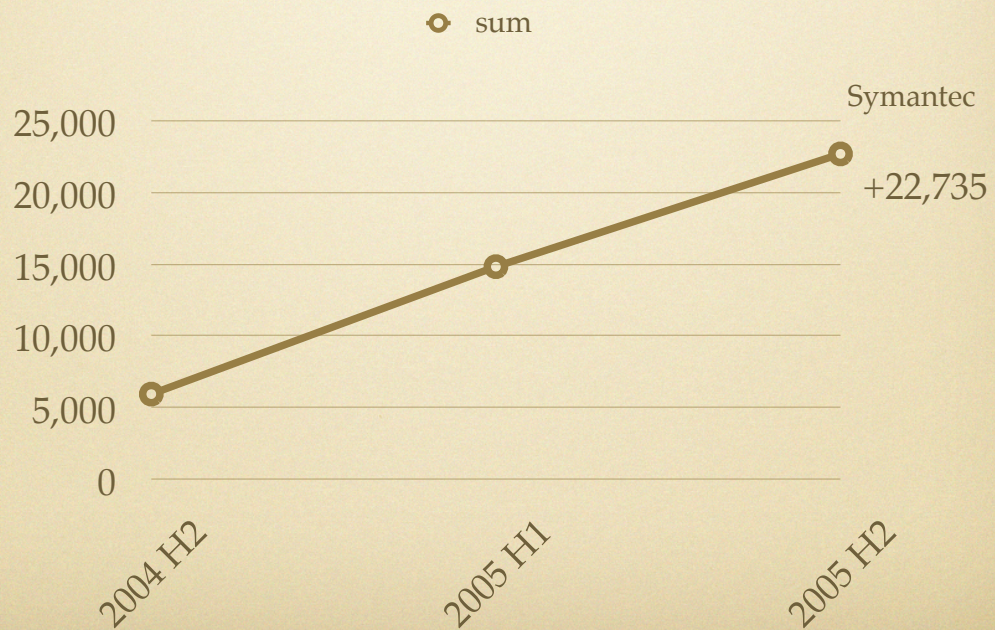


# Bots, top 3 variants



Symantec's semi-annual threat report suggests, as shown here, that there are an awful lot of new variants if existing bots. Note the numbers: that is new variants per half-year. In the case of Spybot, that is 1.5 new variants per hour. This absolutely screams defense-in-depth because blocking by name, by signature, by anything but by effect is lost. One can almost consider variation rates like this to be denial of service (DoS) attacks on the computer immune system. For some reason, Symantec stopped releasing counts of this sort (perhaps thinking it is boring or no longer informative).

# Cumulative for top 3



New variants of just these three bots in a year and a half.

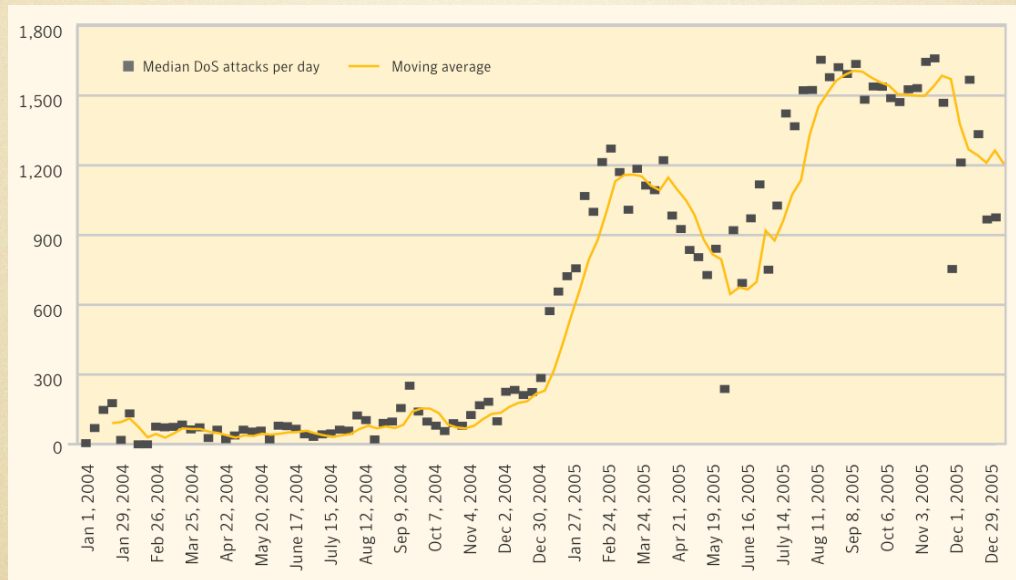
- Net: this is almost impossible to do right, but as with other graphs the trends are not looking good despite the considerable efforts of many people and companies
- Caveat: polymorphism is extremely hard to get a handle on; it is however apparent that some automation is at work hence trends are more a reflection of automation than of level of effort

This is hard to do right in that, largely, we don't yet have long enough tails on the observed distributions and, of course, observing that which does not wish to be observed tends to produce underestimates. At the same time, polymorphism is a growing issue of a very real sort in that it is almost surely now automated if not automatic. Even is merely automated, the attacker has the edge absent a defense in depth strategy on the defender's part.

# Denial of Service

# DoS attacks/day

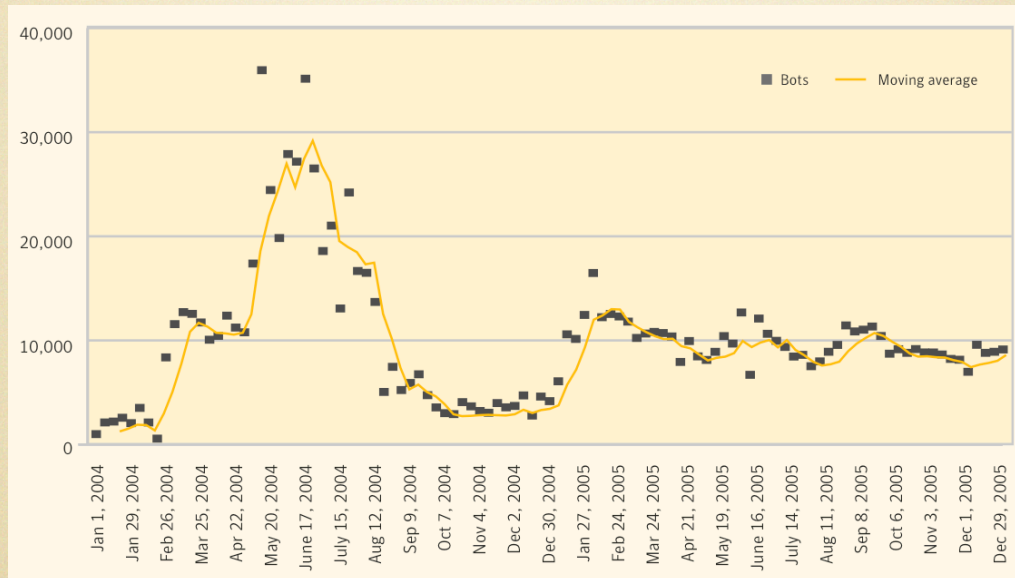
Symantec



Symantec raw data on number of denial of service (DoS) attacks per day. 2005 was not a good year for this.

# Botnet inventory

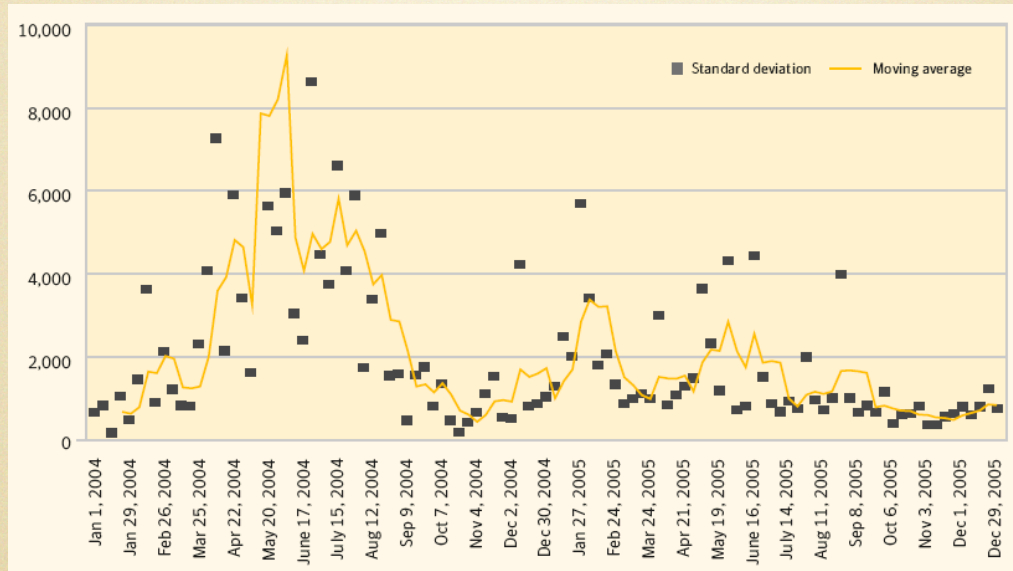
Symantec



This is Symantec's parallel estimate of the number bots outstanding. They suggest a flattening of the curve to circa ten thousand (10,000) at any given time. They do not ask why it would flatten; in my estimation, it flattens because the controllers of the bots have all the prey they can eat -- it is only their satiation that produces this flattening.

# Botnet variance

Symantec



This, however, is interesting. The daily deviation in the number of bots is a measure of the volatility of the bot marketplace and, with high volatility, any supposed flattening of the supply curve says that as fast as bots can be taken out of circulation other bots can be inserted.

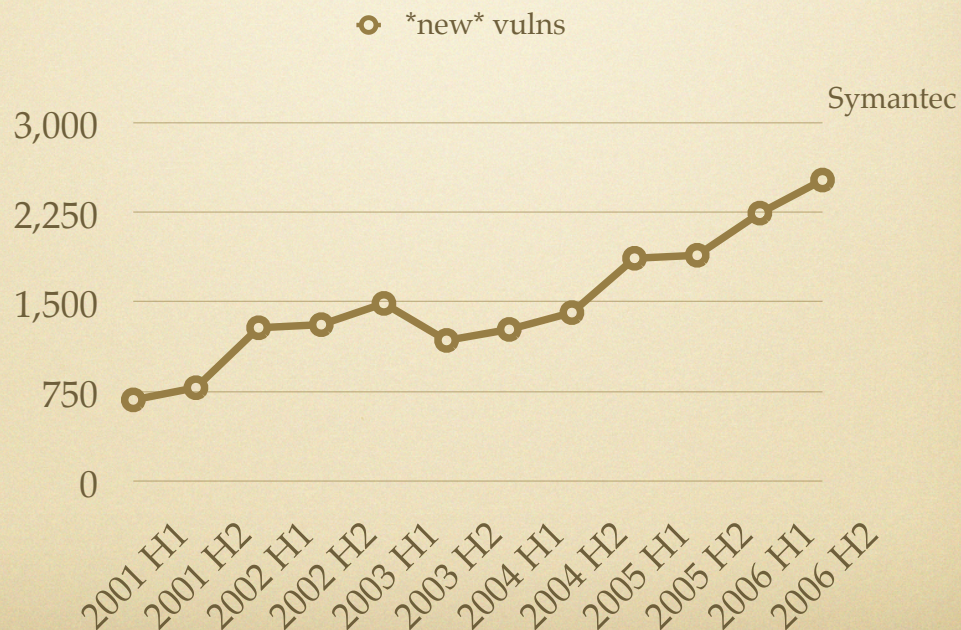
- Net: DoS attacks continue to climb while the number of hosts in botnets appears to be stabilizing; either an efficiency of their use is appearing or the predators are leaving prey on the field
- Caveat: this is probably confounded with the increasing fraction of all attack tools that are themselves stealth, but it also illustrates how hard interpretation of data is

That the inventory of bots remains the same -- just higher volatility -- probably means that the botnet operators have something approximating all they need and just replace repaired machines with new ones. At the same time, increasing occurrences of DoS attacks implies that either the extent of botnets is being progressively ever more underestimated or the botnets are becoming more efficient at doing DoS attacks. Or both. If, as many suspect, the fraction of all attack tools that are stealth is rising, then we may be in for a bad time indeed.



# Vulnerabilities

# Vulnerabilities



There are several sources of this sort of data; this is Symantec's. Note the number is not only multiples per day but rising and it is rising despite a clear effort by all software suppliers to avoid vulnerabilities in the first place. In other words, were it not for the fantastically large amount of effort being put into avoiding vulnerabilities the above curve would be faster rising than it is.

data

685,787,1289,1315,1493,1183,1275,1416,1871,1896,2249,2526

# Cumulative



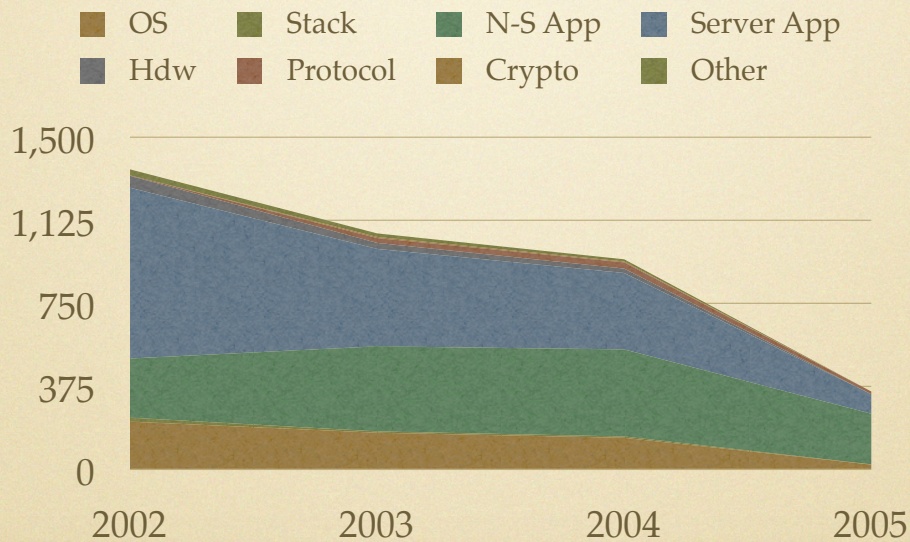
A cumulative picture of the same data. That is nineteen times (23X) more vulnerabilities at the midpoint of 2006 as were present at the start of 2001, again despite markedly higher levels of effort at avoiding them.

# Remote vulns

Component	NIST			
	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

This is data right from the National Institute for Standards and Technologies. I don't like it; it doesn't tell you anything; the column order is reverse chronological and the raw counts offer no insight. But let's start with it, as seen at <http://icat.nist.gov/icat.cfm?function=statistics>

# Overall: progress



Let's see if there is progress being made by making a stacked area graph and running time in the forward direction. It does indeed look like progress.

# non-uniform $\Delta n(\text{vulns})$

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	-na-

-36% CAGR

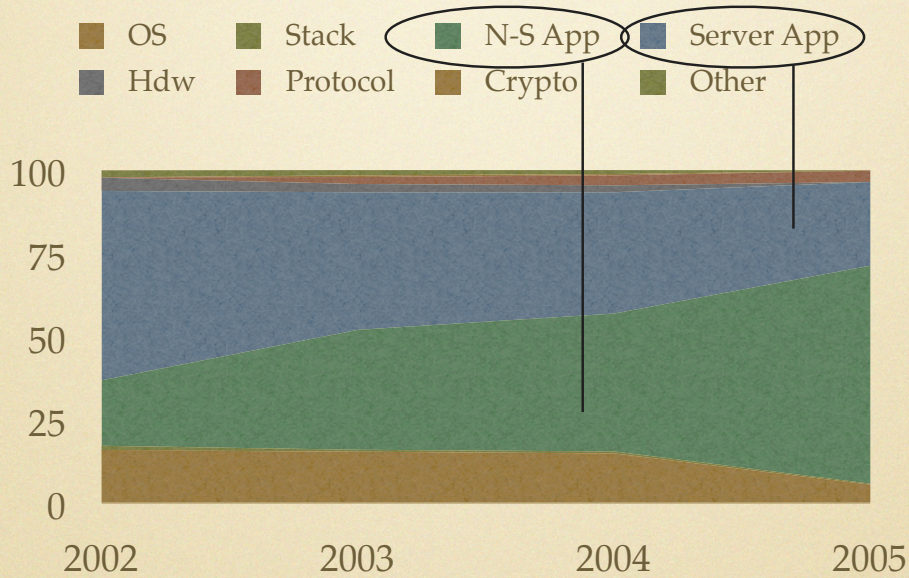
But the progress is hardly uniform. The compound annual growth rate (CAGR) varies from -73.5% to +81.7%, which is quite a range, and has an overall CAGR of -36%.

# Market share

Component	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%

It might be more instructive to look at market share rather than pure count. In the format of the original, it looks like this (which is still pretty useless).

# Δ market share



But as market share we can now see something worth seeing, that the green Server Application category was once dominant but is in fast decline, its place taken by the brown Non-Server Application category.



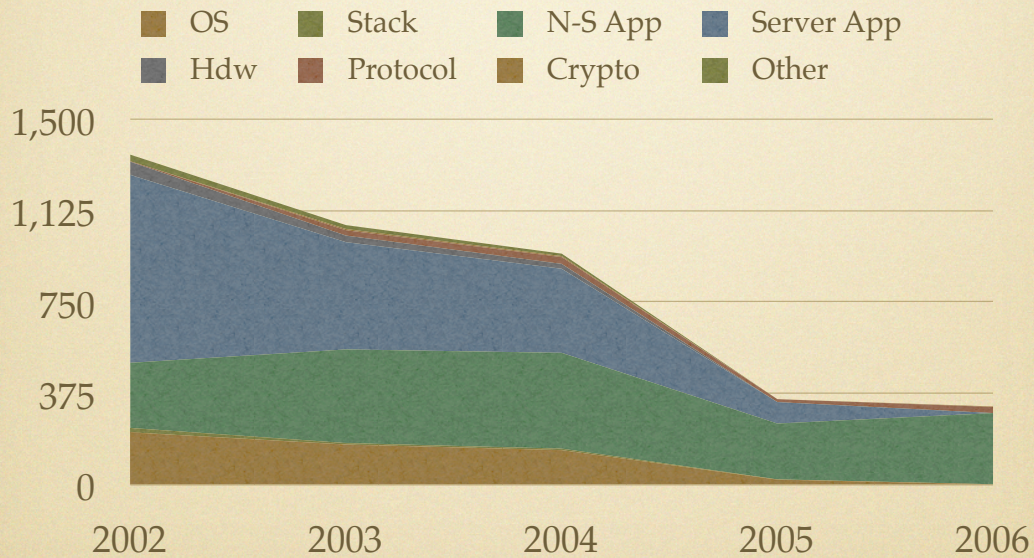
# 2006 forecast

linear regression

OS	0	
Net Stack	0	
Server App	0	
Hardware	0	
Other	0	
Crypto	2	0.6%
Protocol	25	7.8%
Non-Server App	292	91.5%

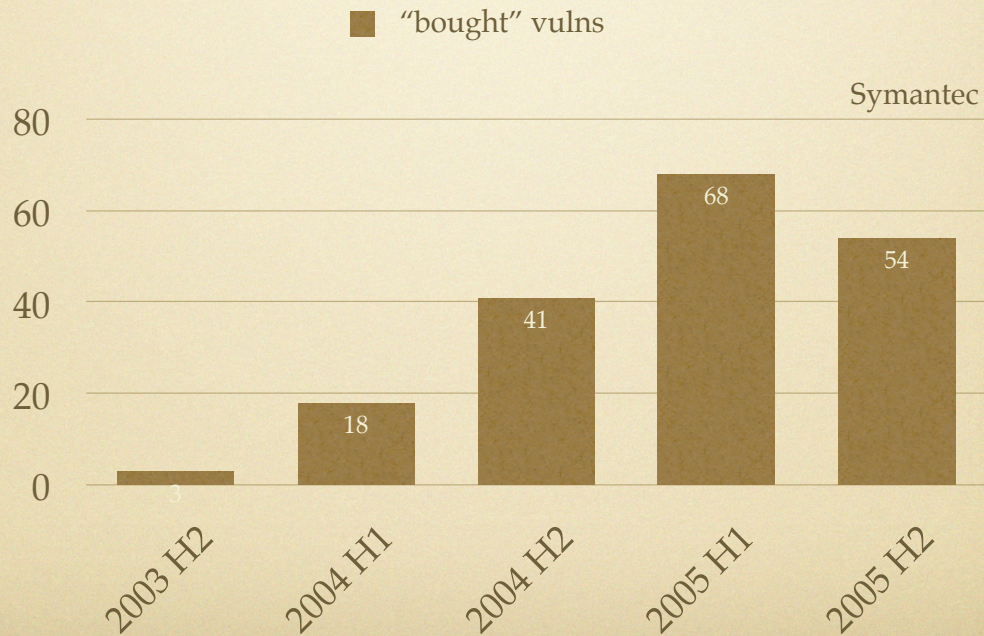
If we take the numbers as given and just do a linear regression so that there is a 2006 (plus one year) prediction, we'd expect the year 2006 values to be down to three (from eight) classes with Non-Server Applications now at 91.5%, thus reinforcing the idea that we need to attend to that line item above all others.

# 2006 forecast



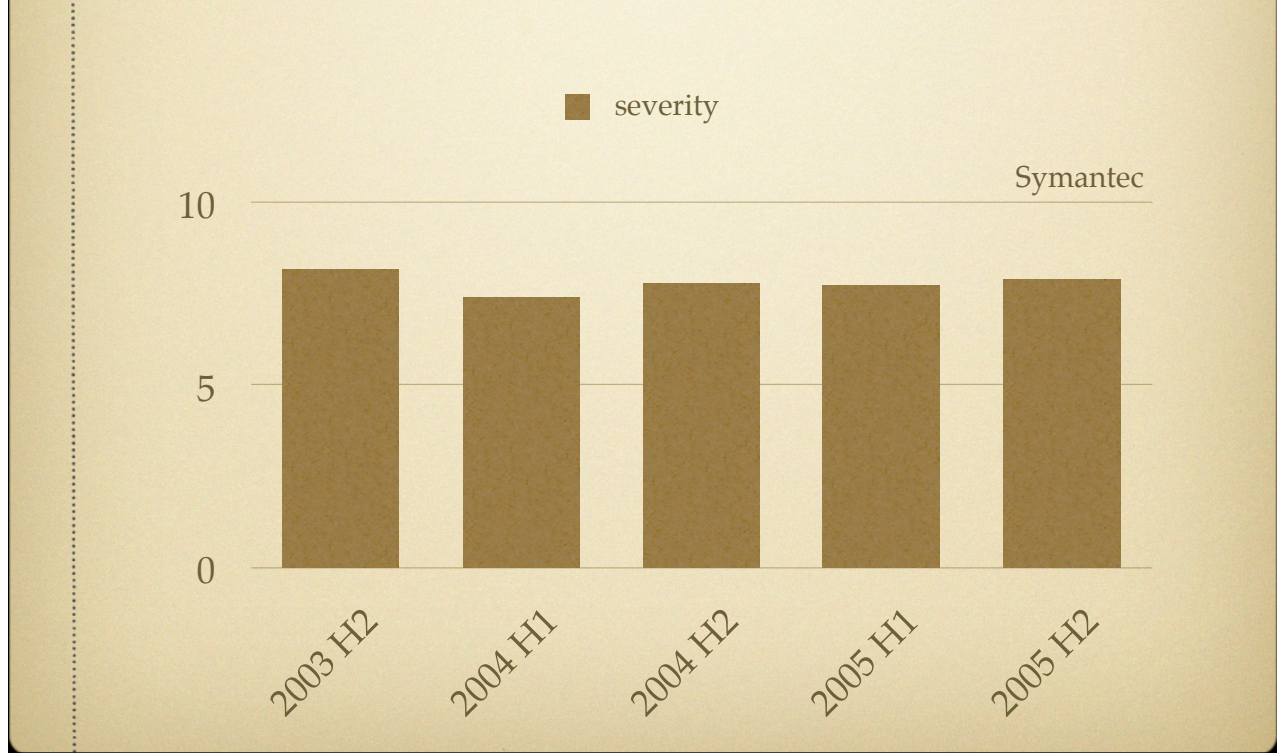
Graphing, in the same style, with the forecast in place gets the point across to almost anyone.

# Open market vulns



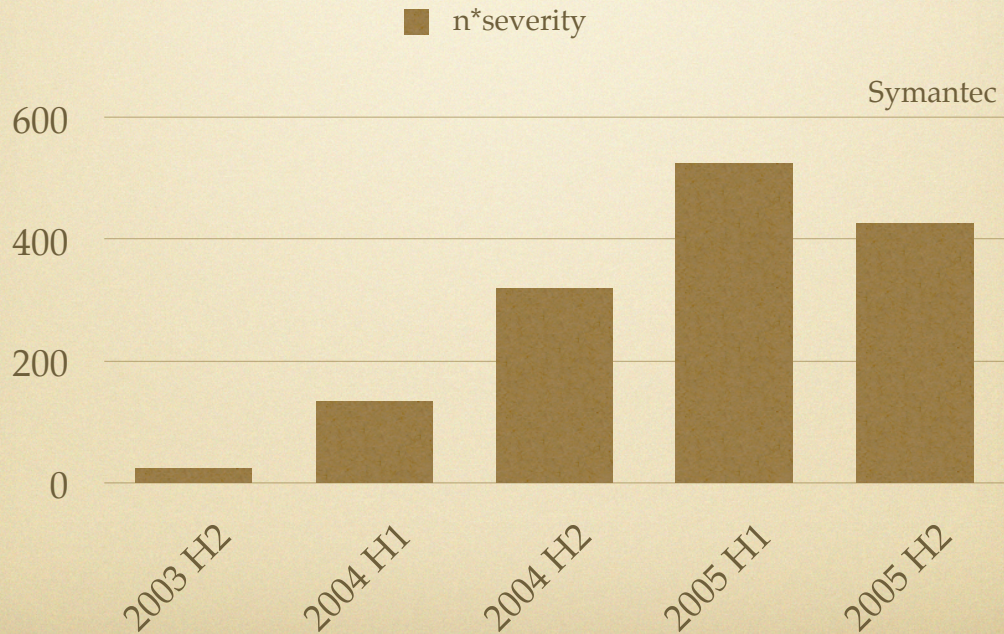
This represents the number of vulnerabilities known to have been bought on the open market. The 2003 H2 figure, not visible on the page, is "3" and that and each of the other numbers in white is how many were bought in that period. Symantec is not a buyer, but does track all known examples of this phenomenon.

# Severity of purchases



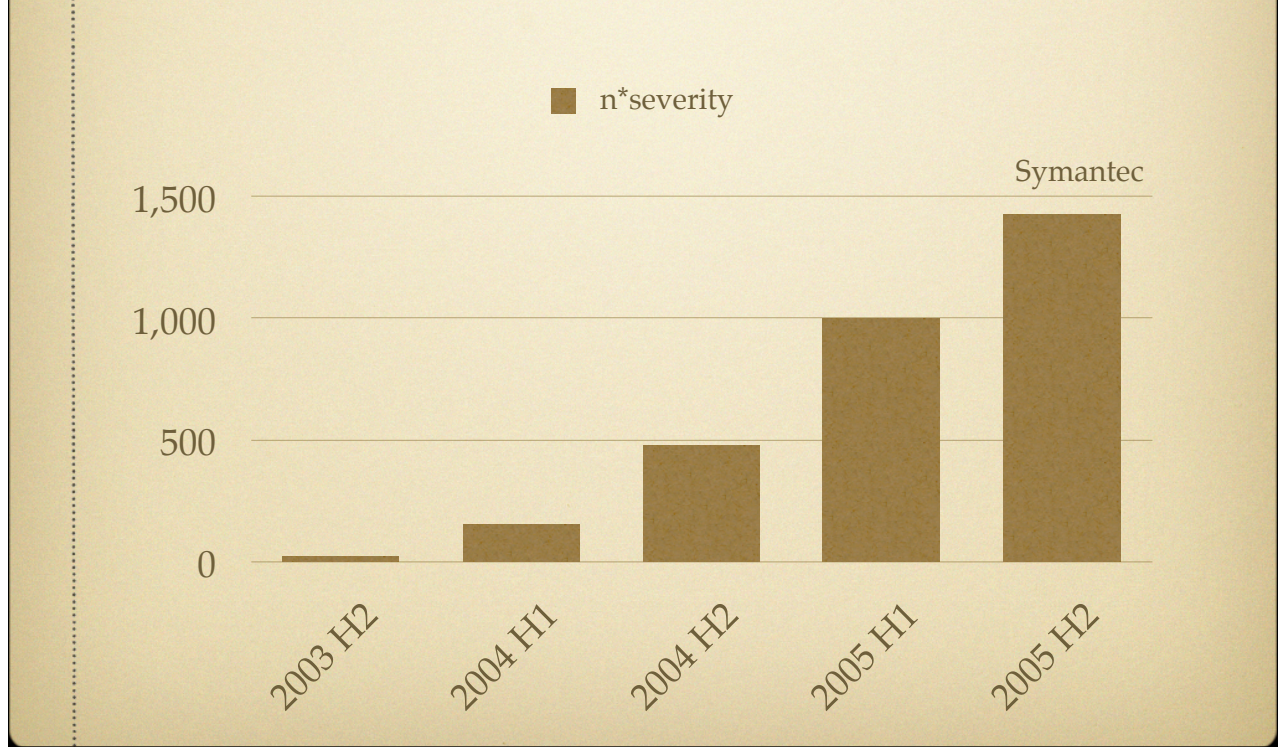
This is less interesting, though it could have certainly been different. What it shows is a constant severity to the vulnerabilities bought (and announced as having been bought).

# Impact of purchases



Multiplying frequency times severity, given a near-constant severity, doesn't actually add much to our understanding though doing this sort of composition is a good thing to do in an exploratory data analysis setting.

# Cumulative



And then cumulating the frequency times severity figures we come to a sense of how strong the current we are swimming in has, in fact, been. As with other Symantec numbers, the source document is their Threat Report IX, March, 2006.

- Net: despite astonishing increases in effort, the number of vulnerabilities continues though the areas in which “remotes” are found has shifted to less important applications
- Caveat: as Symantec says, the decrease may be a mirage if vulns are being held privately at a rate higher than the apparent rate of progress in suppressing vulns in original code

So, we are losing the war but we are losing less slowly than we would be were we not fighting as hard. This would tend to suggest fighting smarter, not harder. In saying that, I am echoing Symantec precisely in suggesting, as they do in print, that “Symantec speculates that while the number of publicly disclosed vulnerabilities could decrease, the window of exposure to potential threats could increase, as details about vulnerabilities are held privately for greater periods of time.”

Spam



# Not just e-mail

- Beginning

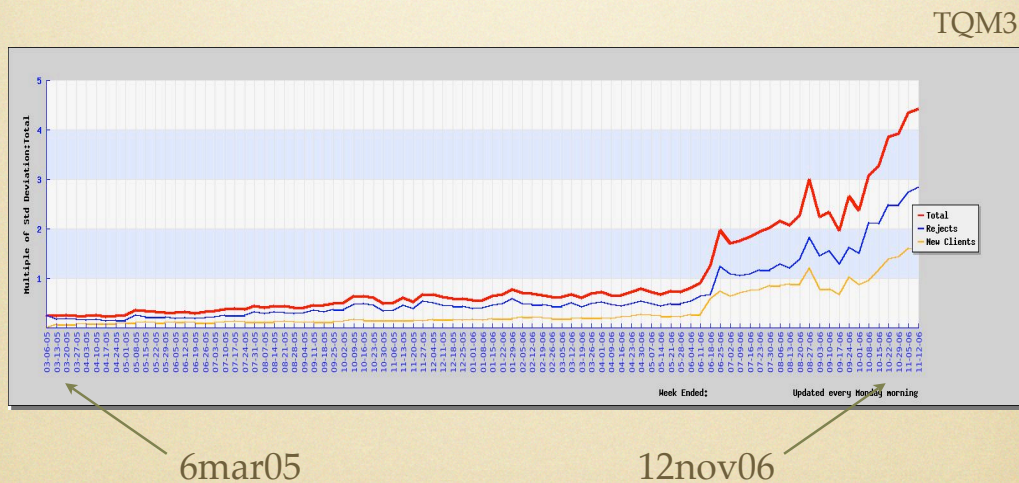
13 April 1994, Laurence Canter and Martha Siegel (Mr.&Mrs.), “green card lottery” legal services, target: USENET

- Latest

Blog-spam; gaming against search engines with keyword-larded RSS feeds

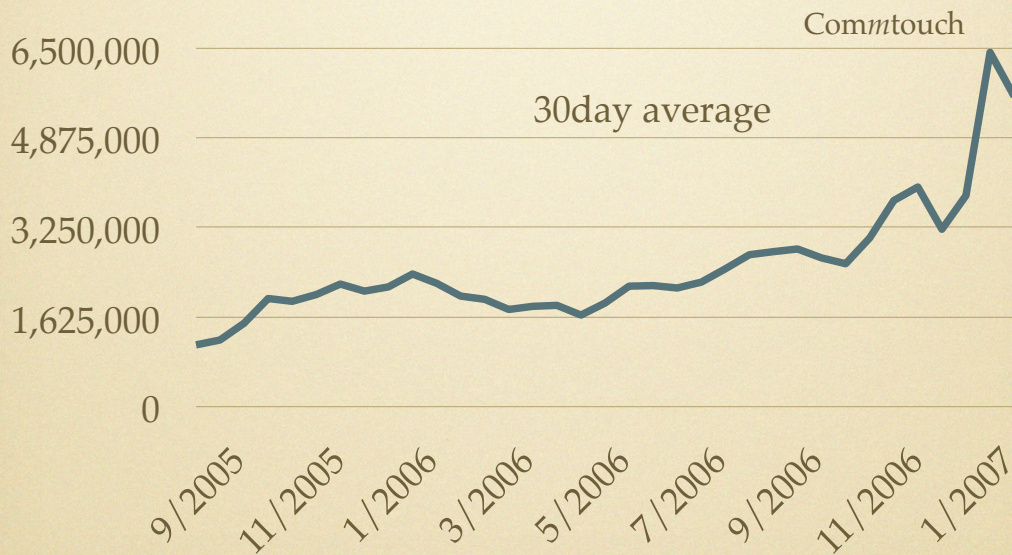
Every channel which does not impose a transmission tariff becomes a channel for unwanted messages. The earliest was USENET, before Canter & Siegel, was largely ruled by netiquette and access was only to noncommercial players. The latest is blogs spewing nonsense RSS streams intending to bias search engines to blog spammers' sites of choice.

# Recent surge



Hard to read, so annotated. The red line is total message count while the blue line is the number of messages rejected as spam. The yellow line is the number of new sender domains seen. For more info, goto <http://tqmcube.com/tide.php>

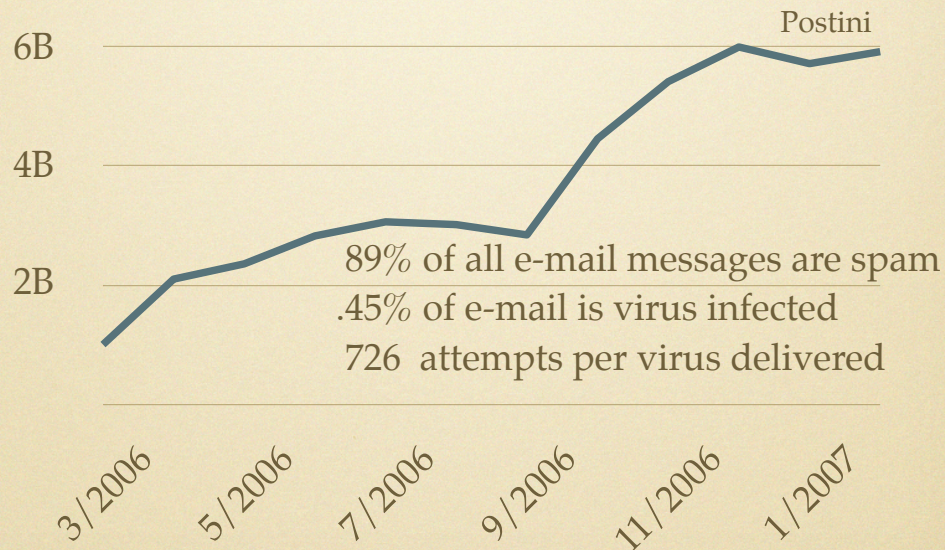
# Corroboration



A different observer, but with similar results modified from raw data for the purpose of this slide by taking a moving 30day average over time. For more info, see (day-by-day data from)

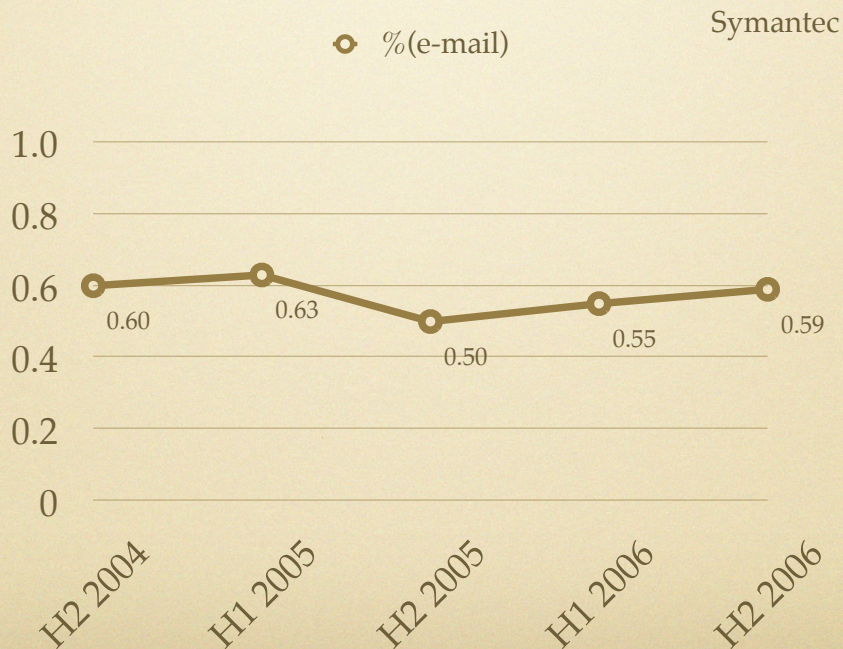
<http://www.commtouch.com/Site/ResearchLab/statistics.asp>

# Corroboration



Yet another observer, but in this case one that both corroborates a recent rise and which, as of 24 Feb 07, gives a 89.0% of all e-mail figure for the spam fraction. They say that their sampling is that 1/371 of all e-mail is virus infected (think defense in depth again) and that every delivered payload cost under 1000 (726) outbound spam e-mails. Compare that to a direct marketing campaign where (a) real money is spent and (b) a 1% return rate is acceptable. Here we have (a) zero money and (b) 45% of an acceptable direct marketing result. In terms of return on investment, the spammer beats the direct marketer cold. See <http://www.postini.com/stats/>

# Different view



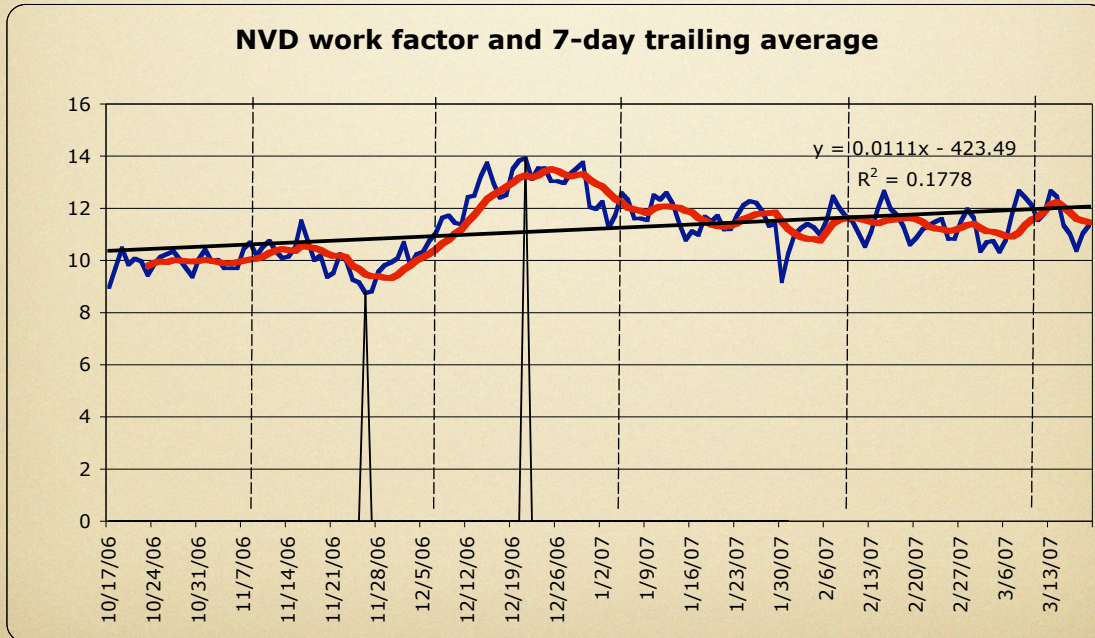
Symantec estimates a constant percentage of all e-mail and lower than Postini, but that is fine and to be expected when you are measuring that which is hard to measure. For Symantec, the information comes from their purchase of BrightMail.

- Net: spam is rising and infecting every new channel as soon as that channel exists
- Caveat: nearly every statistic now has an underestimation bias as filtering mechanisms proliferate both inbound and outbound, especially with major (sue-able) ISPs

Channels are overwhelmed, each new channel in turn. What a surprise. However, as the caveat states, there is an increasing underestimation bias in that the residual spam percentage is after filtering of increasingly vigorous sorts hence that it is as large as it is implies a very high initial transmission rate potential indeed. As the ISPs are beginning to feel heat on this, the filtration has become both inbound and outbound so that, in truth, the “spam rate” might better be expressed as “residual spam rate” since our ability to measure native transmission is probably lost.

Work

# NVD Workfactor

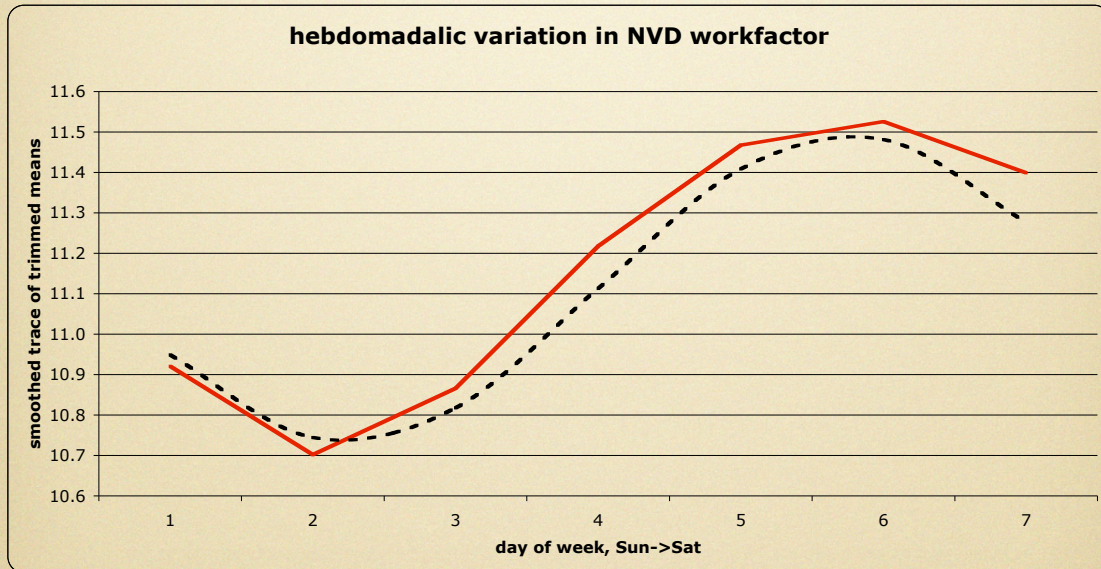


Published everyday at <http://nvd.nist.gov/>, but not otherwise charted. The workfactor number is a composite measure of vulnerabilities and their severities then outstanding.

In this chart, the dotted verticals are Microsoft patch days, the two pyramidal arrows are marking the days of max and min in this window, the blue line is the actual Workfactor Index, and the red line is a moving 7 day average of the workfactor.



# Cyclic, apparently

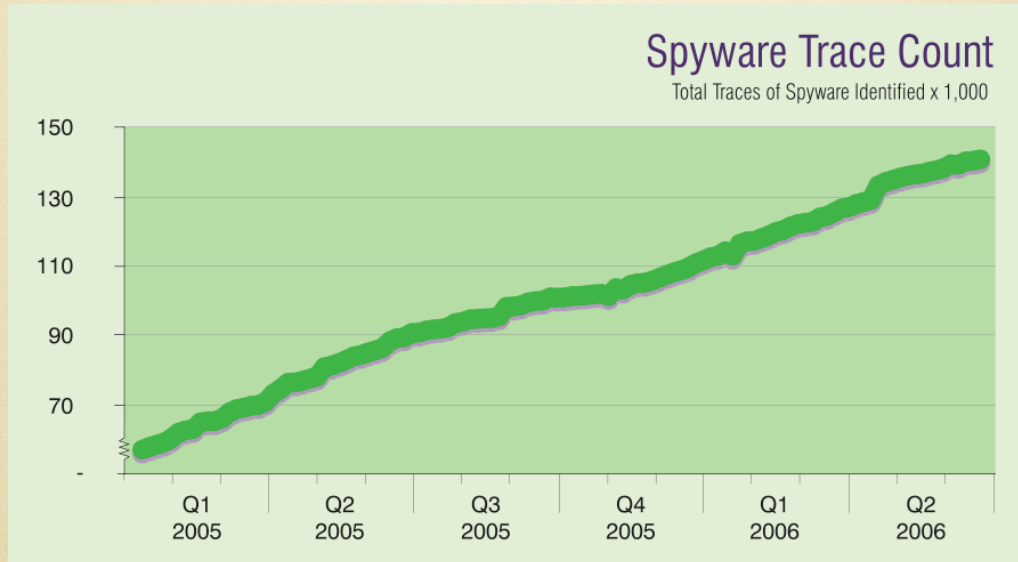


As data accumulates, the curve does reshape from time to time. At the time of this writing, the dotted black line is a fitted sine curve while the red solid line is the mean workfactor by day of the week for the past 120 days.

Spyware

# Steady growth

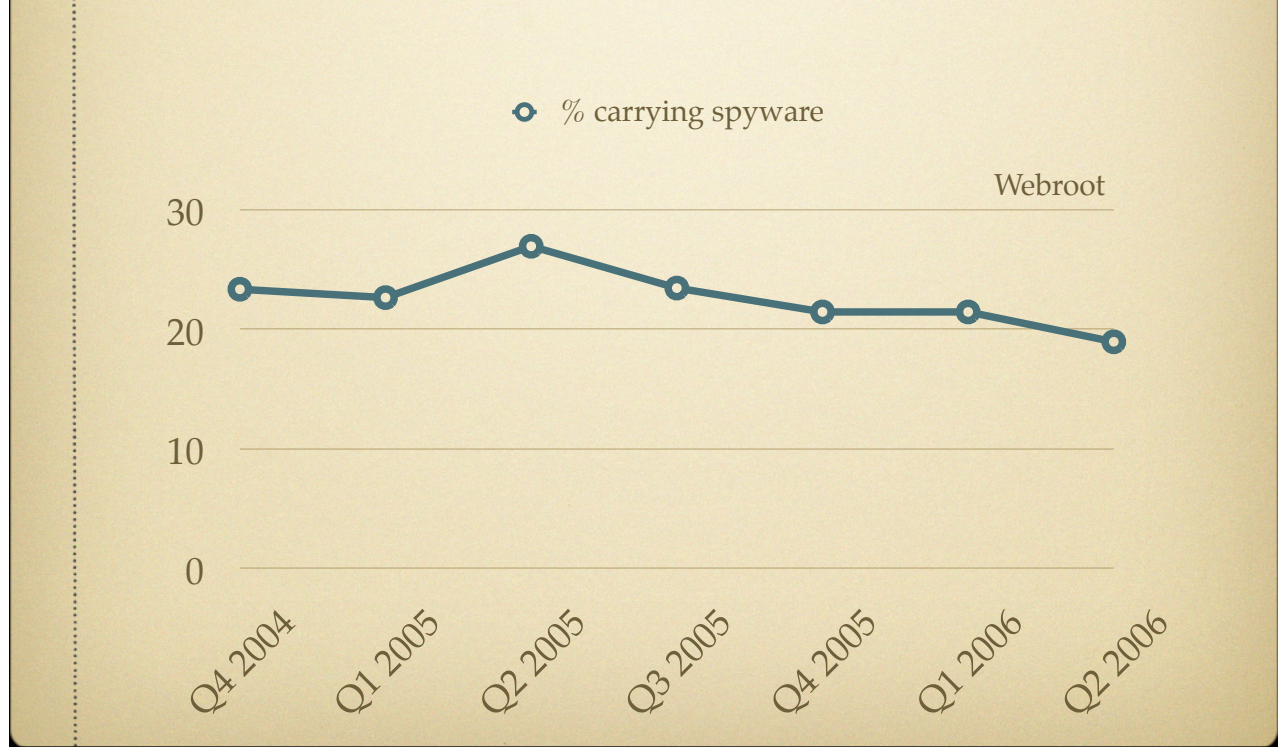
Webroot



42% from China, 17% from U.S.

Note that this graph is cut off at about the 60 x 1,000 mark, but that is for effect (and a commonplace technique in, say, newspaper graphs). In any case the trend is steadily up over five quarters and, from the same report, is currently at 42% Chinese sourced and 17% U.S. sourced. Everyone else adds up to the remaining 40% collectively. See <http://www.webroot.com/pdf/2006-q2-sos-US.pdf>

# Enterprise PCs

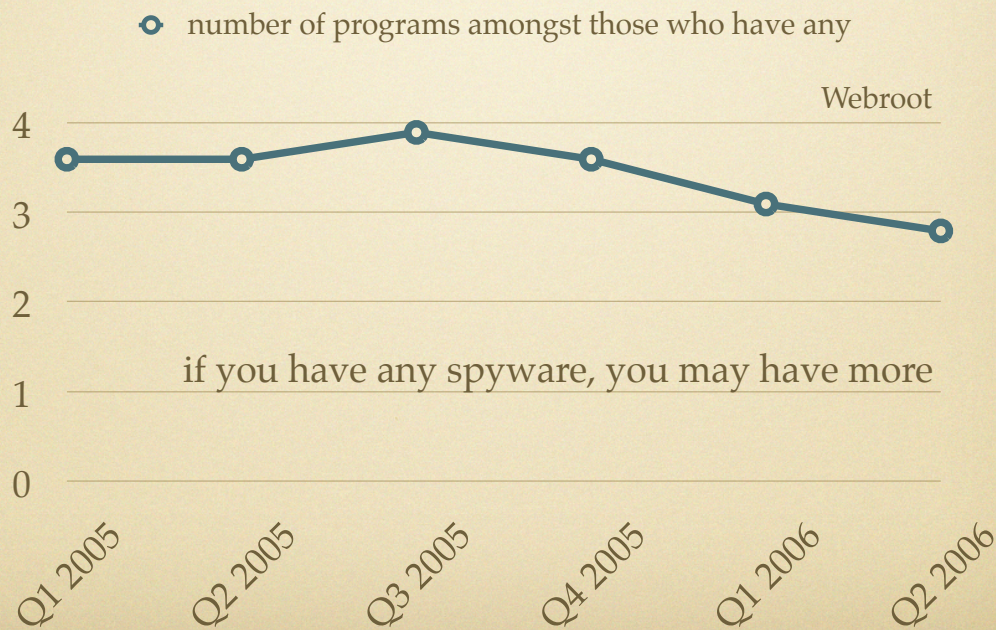


Now, looking at that Webroot data they are estimating a relatively constant 20% of all enterprise PCs contain spyware of one sort or another. Microsoft (not shown) says the figure for “unwanted software” is 67%. Between the two is a good guess but, as seen above and elsewhere, the increasingly strenuous prevent efforts are not yielding a declining infection percentage. If this really is a standoff, then the makers of spyware are able to keep their availability of infected machines constant.

data

23.4,22.7,27,23.5,21.5,21.5,19

# Enterprise PCs

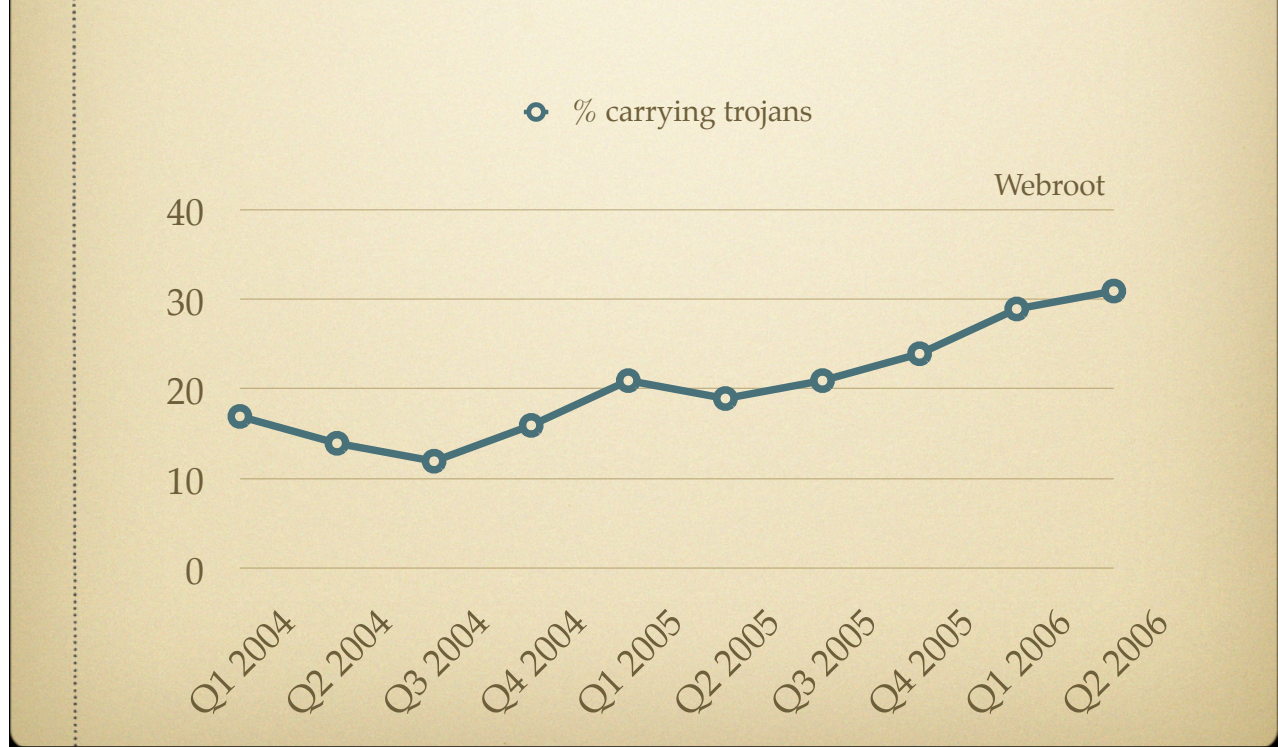


This is interesting, and again an approximate constant, but if whatever it is that you do that causes you to get spyware the odds are that you'll get more than one. Sort of like sexually transmitted diseases, what?

data

3.6,3.6,3.9,3.6,3.1,2.8

# Enterprise PCs

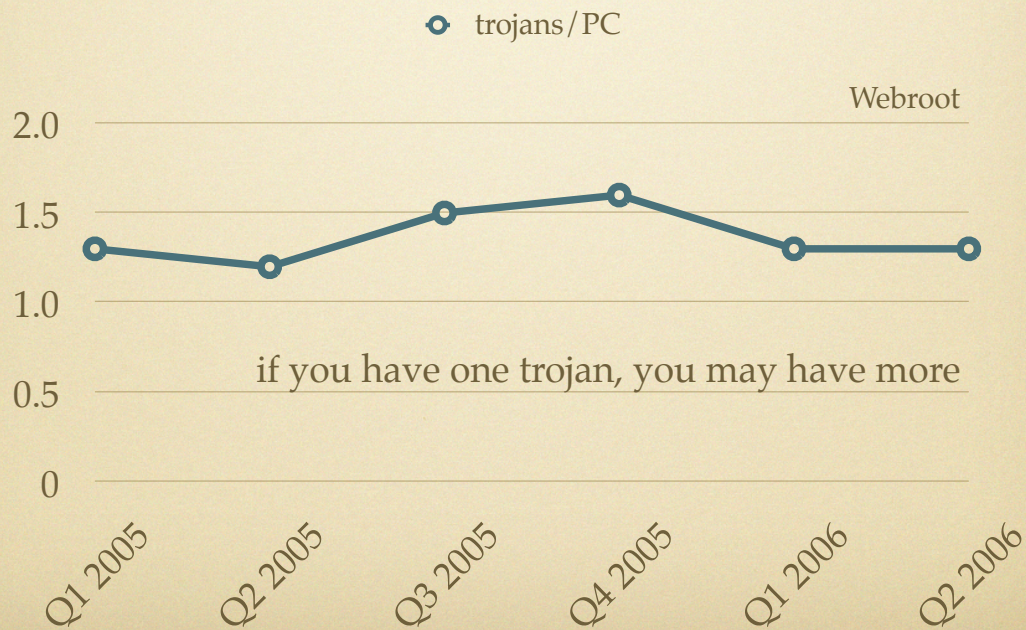


The percentage of hosts that have trojan software. These are both stunning numbers and stunning trends. Amongst other things, it almost absolutely says that a program's apparent name, or pathname, cannot be trusted when making an "is this a trustworthy program" decision, that only some sort of checksum or signature will do.

data

17,14,12,16,21,19,21,24,29,31

# Enterprise PCs

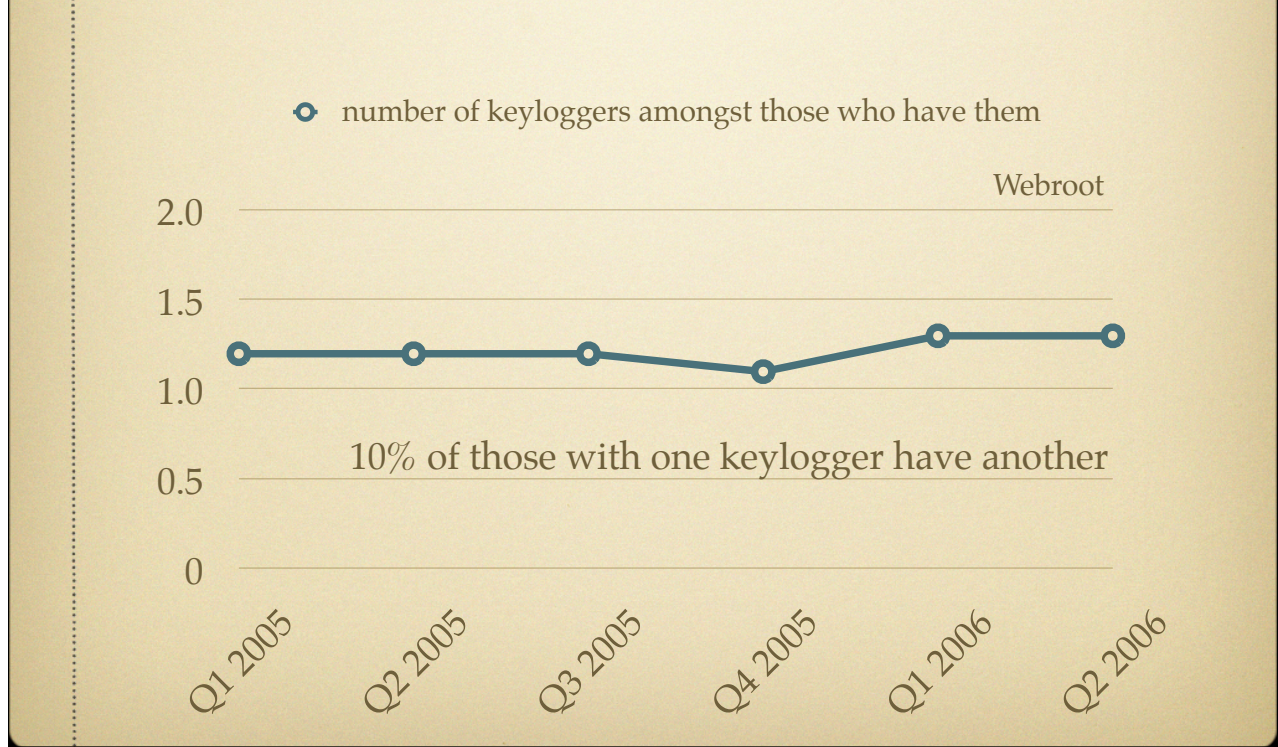


As with spyware, if whatever it is you are doing causes you to absorb one trojan horse program the odds are that you will do it again.

data

1.3,1.2,1.5,1.6,1.3,1.3

# Enterprise PCs



This even extends to keyloggers; if whatever you are doing gets you one key logger the odds are you will get another.

data

1.2,1.2,1.2,1.1,1.3,1.3



*“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.*

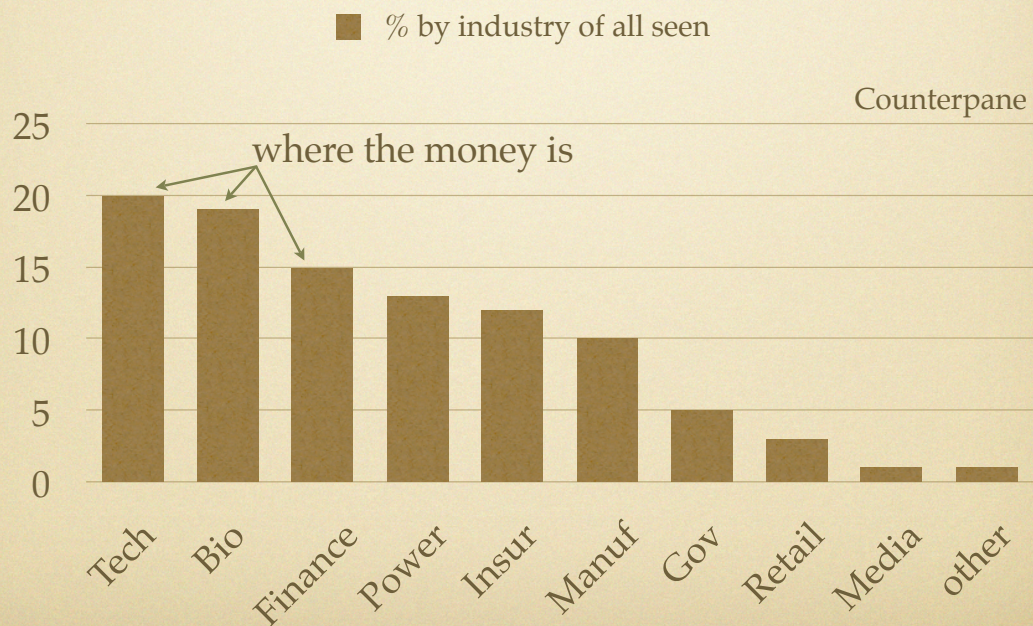
The point here is that one cannot eliminate malware once in place and thus you must either prevent it being put in place or prevent it from taking undesirable actions. This is defense in depth, all over again. See <http://www.eweek.com/article2/0,1895,1945808,00.asp>

- Net: add in that Microsoft says that 2/3rds of all PCs have “unwanted software” and you can see that protection of that data which is software is a more serious problem than any other
- Caveat: definitions matter and the rate at which methods come and go is too fast to develop strongly predictive trend models

Here is a challenge: find trend data that tells you when, not if, protection of data which is software becomes a more serious problem than protection of data which is data in the conventional sense. It is soon, perhaps soon enough that, like the future, it is already here for some enterprises in some areas. As the caveat says, all of this is hard to do when you cannot develop an actuarial tail due to rapid change in terminology or methodology. Not impossible, but not easy.

Who's targetted?

# Policy violations

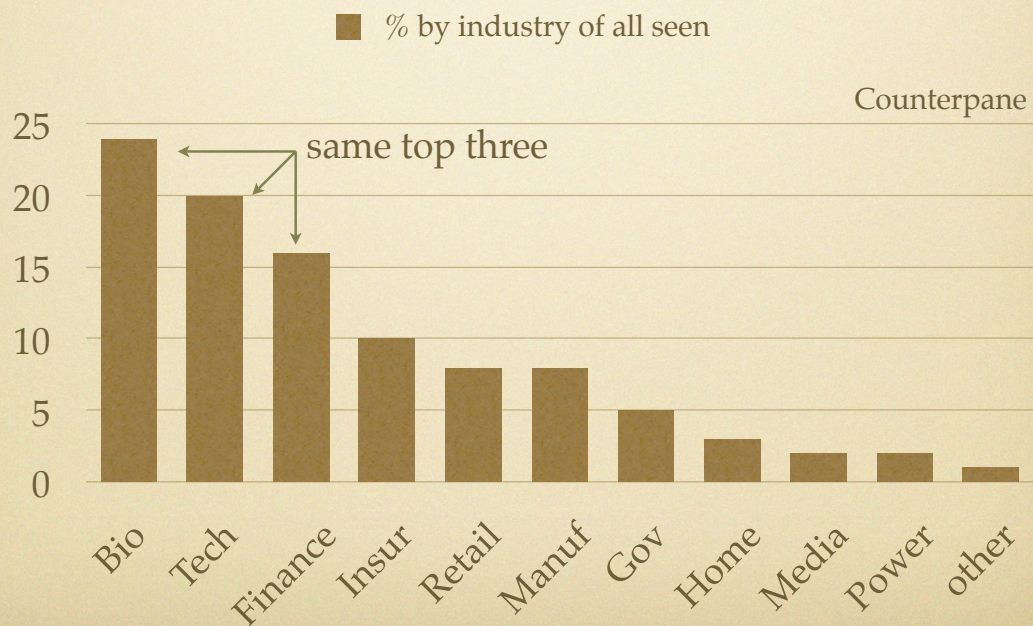


Counterpane Systems is perhaps the original managed security service provider. They are in an observational position for many, many firms and they can thus pool customer data to derive composite numbers that do not expose any individual customer. Here we have their entire mix (summing to 100%) for policy violations and within that mix what industry is the most represented, the next most, and so forth. See <http://www.counterpane.com/cgi-bin/attack-trends4.cgi>

data

20,19,15,13,12,10,5,3,1,1

# Attacks against



External attacks, surprise, surprise, are against the same top three. This is about data as money.

data

24,20,16,10,8,8,5,3,2,2,1

- Net: in a world where gaining money has supplanted gaining notoreity as primary motivator, targets are predictable
- Caveat: non-trend prevalence data only, and only from one monitoring firm (which, to its credit, does publish)

Though the Counterpane data were not trends in time, they were trends in targetting and they show that where the money is is where the attacks go whether they are outside attacks or internal policy violations. While this is only one firm, it is a form which other MSPs can use should they care to publish their pooled data (and any holder of pooled security data has a professional ethics requirement to do so).

# Public Health

# Detection is doomed

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

*...that's in the large, but in the small...*

Source: Symantec Threat Report IX & XI, March, 2006 & 2007



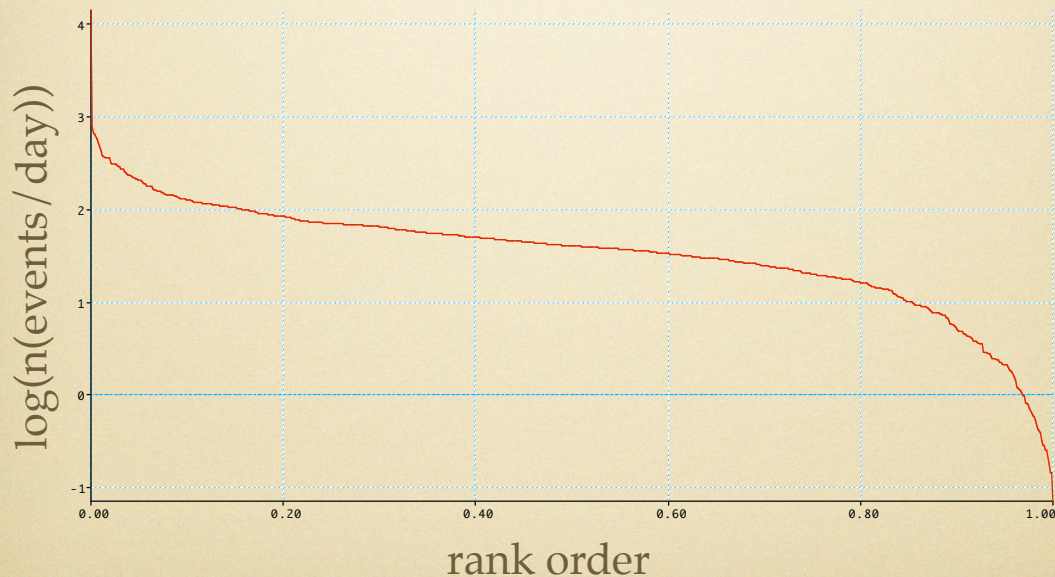
## Data handling violation events highly skewed

location	a single user	50% of all events from
Domestic Sfw House	18.8% of events	5.7% of users
Trading Bank	25.2% of events	6.9% of users
Pro Sports Team	15.9% of events	6.7% of users

Elsewhere in my professional life, I look at data handling policy...

These serve to illustrate what a “data naturalist” might find on three different islands. As you can see, violations of data policy (as recorded by the surveillance software) are predictably skewed -- the many with a few violations and the few with many violations.

# Event skew, same bank



Another way to look at the skew between the few with many violations and the many with few. For the several thousand individuals whose actions were surveilled, this graph rank orders them by number of violations (X axis) and for each person how many violations they had (Y axis, on a log scale).

That the majority of this graph is linear on the log scale and is overall a sigmoid curve is interesting and obviously supports certain models. Outlier handling is thus an important question for the data security manager.

# Domestic Chip Fab



This is the rate of data-handling policy violations at a domestic chip fab, showing a crescendo as the week proceeds. There is no known explanation for this.

# Asian Chip Fab



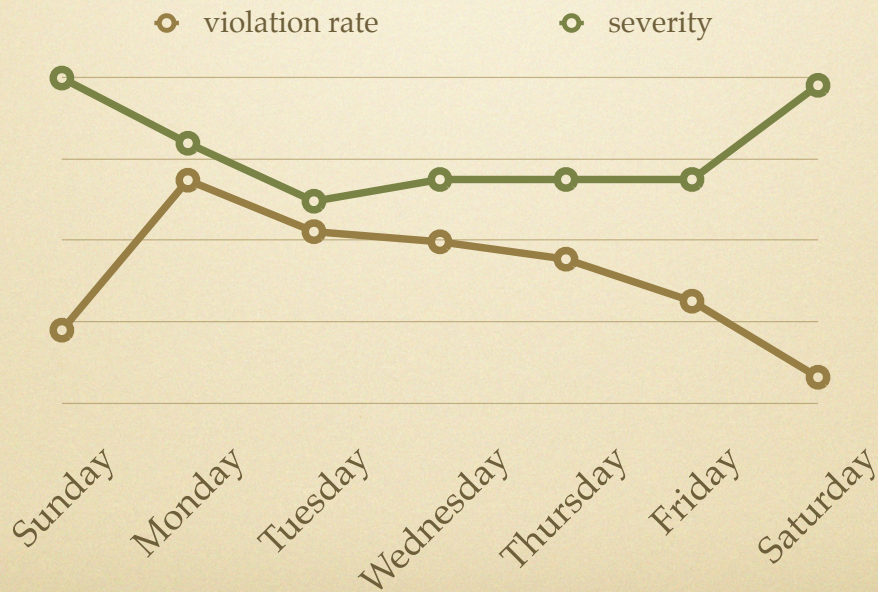
This is a modest Asian chip fab and, in contrast to the domestic chip fab, data policy violations start with a bang and go to zero as the week proceeds. There is no known explanation for this, either.

# Domestic Sfw House



For a domestic software house, there is a different pattern, again with no known explanation. Yes, “no known explanation” can be “nothing but random error” but it is still interesting to look a little deeper -- and is now possible to look a little deeper.

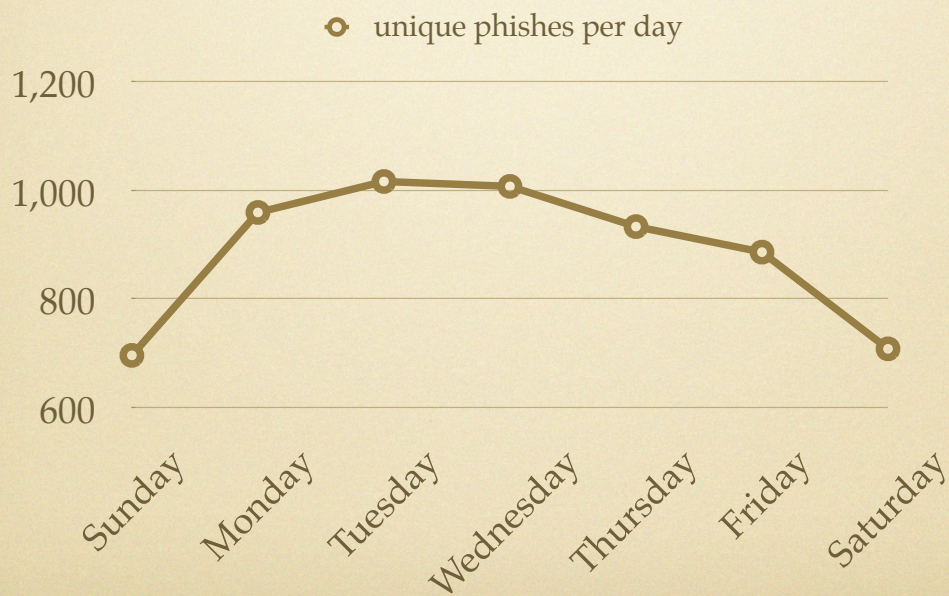
# Trading Bank



This is a (different) domestic trading bank showing an effect that begs for closer inspection in that the rate of violations seems to be approximately inverse to their severity.

It's easy to make behavioral hypotheses when you see data like this.

# Phishing is a job



From Symantec ISTR:xi, the living proof that phishers are working a job.

Where this leads



# Security & threat co-evolve

- Predators are the reason prey evolve & diversify
- Virulence is eventually proportional to immunity
- Infectious agents rarely cross species boundaries
- Corruption of immune system worst (Witty)
- Parasites co-exist non-lethally (botnets)
- Evolution's course is by punctuated equilibria

Nature has a lot to teach us, and all of this was learned from a bacteriologist who is in turn looking at computer viruses as life forms.

# Internet as punctured equilibrium

- Irresistible economics force participation
- Location independence of prey and predator
- Force multiplication  $\propto$  bandwidth
  - Bandwidth is cheap, especially if you steal it
- Economic driver for commoditization
  - Hence monoculture and monoculture threat

The Internet punctured the then equilibrium for sure, and in particular one must participate in it but that increases target density for the opponents. Worse, as a commodity it produces the monoculture we see around us and a monoculture is a public health disaster.

# Data has value

- Growing fraction of total corporate wealth
- Growth in data volume parallels growth in the value of corporate data
  - Installed capacity: +150% / annum
  - Total retained volume: doubling every 30 months
- Magnitude = data volume  $\times$  unit value  
Sign = + if used well, - if not

Data has value. The numbers come from private reports by Forrester and Gartner. The comment at the end is to remind you that value is really  $\pm$ value.

# Data is mobile

- The optimal computer will change:
  - cpu/disk/bandwidth doubling at 18/12/9
  - 10 year implication: 10 times more mobile though 10 times more data per unit of CPU
- Winners will have the most information in play  
Losers will have too much
- Convergence of pure comms (telephony) and data rich applications

Data is mobile, and in particular it becomes more voluminous enormously fast but mobile faster still. The nature of the optimal computational layout changes, and probably back towards what might be called “time share.”

# Data is now focus

- Security is what distinguishes data which has value from data which does not
- Rising threat requires any defensive perimeter to contract
  - True for the military, wildebeeste, and data
- Contracted perimeter for data shifts focus to individual data objects at their point of use

Therefore data is our focus as security people going forward, and as the threat is rising our perimeter is contracting. This is a universal truth that just happens to be playing out with data as well.

# Operationally

- Data is at risk when it changes from at rest to in motion
  - Think state-change (like evaporation)
  - Point-of-use  $\equiv$  where that state change occurs
- Monitoring is the first priority
  - You cannot control what you cannot see
  - The unknown unknowns will kill you; Donald Rumsfeld was right on
- Security metrics therefore begin with certainty at the point-of-use

We have to, therefore, monitor the place where data changes from at rest to in motion, and until we do we are disabled in precisely the way that Don Rumsfeld meant: We'll be faced with unknown unknowns.

# Losing propositions

- Content inspection – Can be defeated by Pig Latin, much less encryption
- Statistical anomaly detection – Infeasible work factor to damp out false positives
- Signature finding – Red Queen: "...it takes all the running you can do, to keep in the same place"

These are all wrong, if not this moment soon in the future. Content inspection is trivial to defeat when you know it is there, anomaly detection cannot catch everything without drowning you in data, and the anti-virus signature problem is already a failure.

# Problem Statement

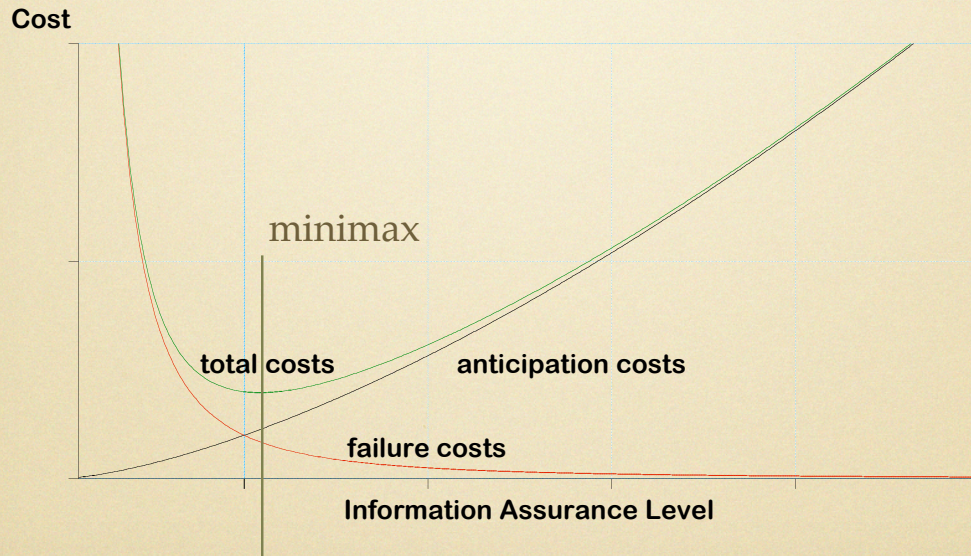
- Data protection that is
  - Inescapable
  - Invisible
  - Future proof
- With optimization between
  - Anticipation costs (preventing trouble)
  - Failure costs (cleaning up trouble)

This is where we have to go, and it is all about the data.



# Bear v. Avoid

NCMS



Risk transfer is about trading one risk for another; that can be internal as well as external. This picture does not specify, but it illustrates the tradeoff between anticipation (prevention) costs and failures (mitigation) costs. The total cost is the sum of the two and, as the graph shows, spending nothing on anticipation maximizes failures costs just as spending too much on anticipation minimizes failure costs. The saddle point is your management target.

National Center for Manufacturing Sciences, August, 2002;  
<http://trust.ncms.org/pdf/CostInfoAssur-NCMS.pdf>

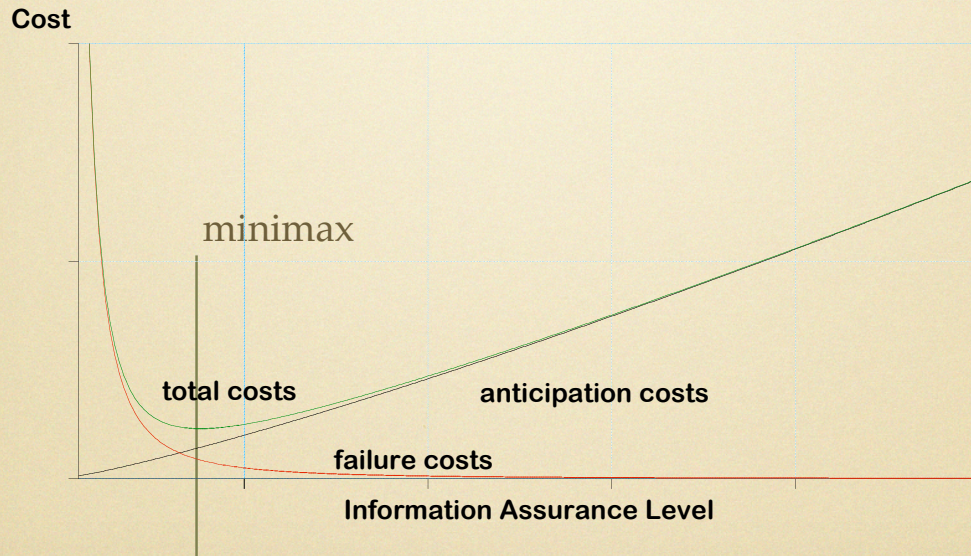
# Setting matters

- Little collaboration  
⇒ low failure cost ⇒ spend little
- High collaboration  
⇒ high failure cost ⇒ spend more

One of the things that NCMS points out well is that the level of collaboration you have with your customers, suppliers, and other counterparties affects the cost of failure should you be unable to have that collaboration. If you have little collaboration, you can be offline, say, at little effect. If you have a high degree of collaboration, the effects of being offline are more profound. Were these true, you might have to adjust your spend up or down to reach optimality.

# Lower collaboration

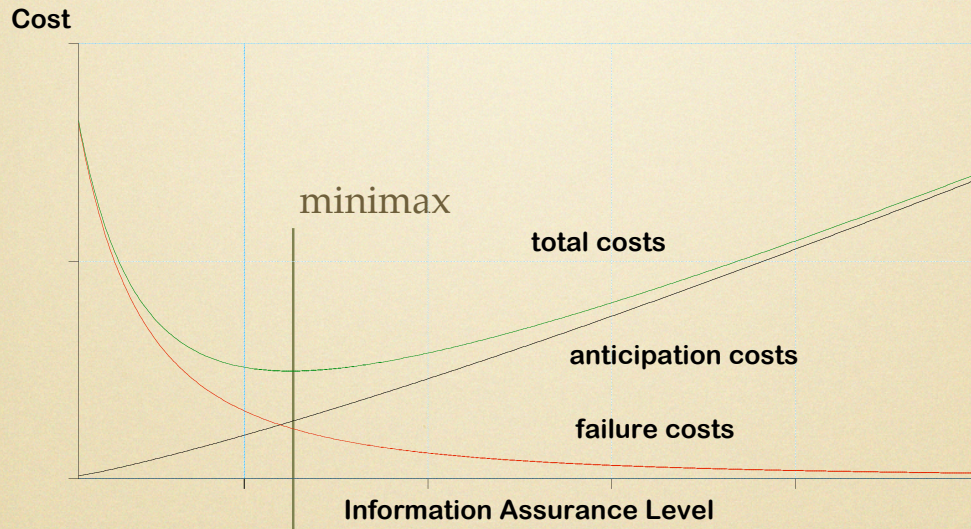
NCMS



So at low collaboration, the total cost has its minimax point where anticipation costs are minimal because failure costs are also minimal.

# Middling collaboration

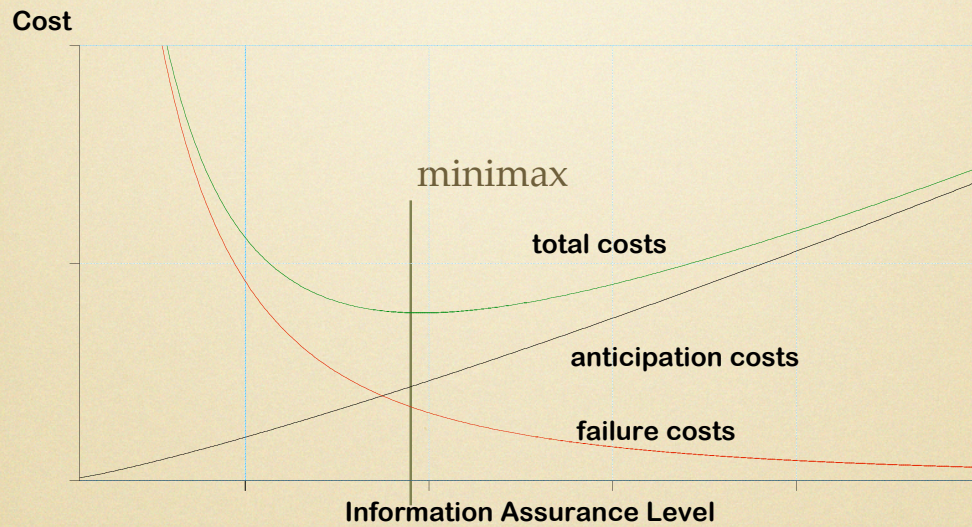
NCMS



At middling collaboration, the failure costs have risen so the minimax point has moved rightward.

# Higher collaboration

NCMS



At high collaboration, more money still must be spent on anticipation if the minimax point is to be achieved.

# The fork in the road

Non-revelatory attacks require preemption,  
...and preemption requires intelligence,  
...and intelligence requires surveillance.

Which is your unit of observation?

- One person
- One data item

So if we are to get preemption for which we need intelligence that comes from surveillance, what is the focus of our surveillance? Do we instrument the data or the people?

# Corroboration

Jeffrey Ritter, Esq., on today's legal reality:

That which is not documented does not exist.

That which was not recorded did not happen.

That which has not been audited is vulnerable.

He does not mean a path to invisibility, but rather that these are the pre-conditions for liability. He is advising law firms on just this sort of thing, i.e., that their own handling of co-mingled documents from their clients is dangerous to their clients and themselves unless that handling is done with rigor. (His firm is Waters Edge Consulting, [wec-llc.com](http://wec-llc.com), co-founded with Karen Worstell, former CISO for Microsoft.)

In the end, they will their freedom at our feet  
and say, "Make us your slaves, but feed us."

"The Grand Inquisitor," The Brothers  
Karamazov, Fyodor Dostoyevsky





The time is up. The problem is hardly tasted, much less solved.

Feel free to contact for more (there's lots): Dan Geer,  
dan@geer.org, +1.617.492.6814