White Paper

Security of Information When Economics Matters

VERDASYS™

Daniel E. Geer, Jr., ScD
Vice President, Chief Scientist
Verdasys, Inc.

May 2004

ABSTRACT

Establishing, growing and protecting competitive advantage is the responsibility of a management team at the highest levels. When that advantage is directly related to the information you have – about your customers, your market, your future plans – executives must explore sound economic models for putting the security of information in place. We will discuss the security of information as an economic matter using cost-effectiveness as a decision support tool. We will distinguish between those information security costs that are made in anticipation of preventing problems and those that are made in response to information security failures that have already occurred. We will argue that the trends are discouraging and the money being spent needs to be strategically redirected. We will say where that redirection should be.

Information security provides the maximum protection at the minimum cost.

INTRODUCTION

For security technologies and security practices to have any long term hope of adding value to the enterprise they have to be both expressed in rational economic terms and subject to economic tradeoffs. We have to know what we are protecting and we have to protect what we know to protect. There are thus two wings to this discussion: What is the rational valuation of that which is to be protected and what, differentially, are the available means to do so. This is not where we are; rather we have to start from where in fact we are today.

Premise:

- In any stable country, economics matter most.
- In any advanced country, information is an increasing share of the intangible property on which that country's advanced state is based.

Consequence:

- In any stable and advanced country, the economics of information security are more crucial with every passing day.

Because the examination of information security in the light of economics is relatively undeveloped, our arguments will be definitive where they can be but indicative otherwise. Information security is in somewhat desperate need of risk analytics that are well developed enough to steer business in the way it likes to be steered, which is to say to "follow the money." While it is early in the game to be putting words on paper, somebody has to go first.

The term "information security" has some meaning as a set of operational practices. We are reluctant to fiddle with the meaning of standard definitions of any sort because 99 times out of 100 fiddling with definitions misleads novices and confuses practitioners, but information security does not yet have consensus precision. Let's just say that for the purpose of this paper we are stressing one narrow definition of information security:

Information security provides the maximum protection of information at the minimum cost.

This is a so-called "minimax" solution to treating information security as a resource (budget) allocation problem. If allowable cost is fixed in advance, such as by budget, then this is a classic economic optimization problem.

APPROACHES TO ANALYSIS

There are two well-known, well-established ways to look at this minimax question: As a problem in cost-benefit or as a problem in cost-effectiveness. So-called "cost-benefit analysis" evaluates the ratio of benefits to costs. It is appealingly rational and it is difficult. It is difficult precisely because it requires costs and benefits to be calibrated in some common currency: dollars or units of pain or something else in which the costs and the benefits can both be expressed.

To illustrate the difficulty of cost-benefit calculations, let's look to health and health care: How many US dollars is a human life worth? How many dollars is it worth spending to save a human life? When is spending more dollars no longer wise because the lives saved are too expensive to save? Too expensive for whom?

The above is not to start a fight; it is to illustrate that when doing cost-benefit analysis you have to be able to equate the cost and the benefit on a single scale, e.g., would you rather have $10M or another year of life? An afternoon to yourself or a month of heating oil? Pervasive wireless access or an improved spam filter? And so forth. Cost-benefit analysis is great when you can do the valuation that converts apples to oranges, but not a good idea otherwise.

The alternative to cost-benefit analysis is cost-effectiveness analysis. Cost-effectiveness escapes the common valuation problem by assuming that you will incur the cost regardless, hence the remaining question is what is the most that you can get for that cost. Cost-effectiveness is to cost-benefit as investment is to gambling – more rational but less exciting. Where cost-benefit asks "Would you rather have $10M or another year of life?" cost-effectiveness asks "If you are already committed to spending $10M would you rather have another year of life for yourself or for each of your children?" Cost-effectiveness is tractable everywhere whereas cost-benefit is tractable only for special cases.

To illustrate, we'll turn again to health care simply because health care is more likely to be broadly familiar. Here are cost-effectiveness figures from the National Center for Policy Analysis:[1]

By spending $182,000 every year for sickle cell screening and treatment for black newborns, we add 769 years collectively to their lives at a cost of only $236 for each year of life saved.

By spending about $253 million per year on heart transplants, we add about 1,600 years to the lives of heart patients at a cost of $158,000 per year of life saved.

Equipping just 3 percent of school buses with seat belts costs about $1.6 million per year; but since this effort will save only one child's life every year; the cost is about $2.8 million per year of life saved.

We spend $2.8 million every year on radionuclide emission control at elemental phosphorus plants (which refine mined phosphorus before it goes to other uses); but since this effort will save at most one life every decade, the cost is $5.4 million per year of life saved.

The analogy with information security is clearly that the cost of prevention is not one-for-one with the value received for those costs. The effectiveness of various health interventions differs over a wide range; security has a similarly wide range of effectiveness.

In first world economies, winners have the most information in play while losers have too much.

In first world economies, winners have the most information in play while losers have too much. Security is that fine line that separates those winners from those losers. As we have

[1] *Dying Too Soon: How Cost-Effectiveness Analysis Can Save Lives*, NCPA, Washington, D.C., 1997, http://www.ncpa.org/studies/s204/s204.html

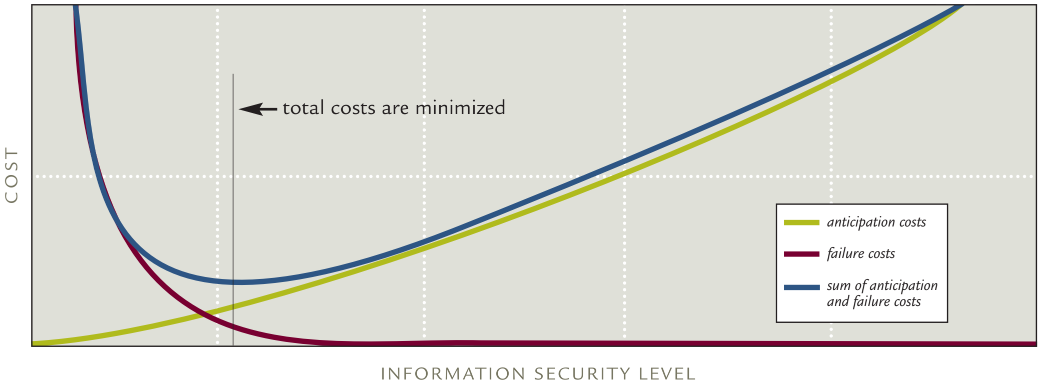
The most influential calibrator for the right amount of security spend is how much collaboration you have.

argued elsewhere[2], protecting individual items of data is where the action is or will be soon. To synopsize that argument, the more open the corporation's business model, the more information security contracts to the point of use, i.e., to the individual data object. For example, your organization is considered to be more "open" if you have a high number of electronic interactions with partners, customers, suppliers and the like. This means a shift in exactly what problem it is that optimal information security is positioned to solve, what an effective solution looks like, and what it costs.

So let's look at the cost and effectiveness of information security. We again have a fork in the road, though in this case we are going to take both of them. Using the terminology of one consensus study around the costs of information assurance[3], one of our forks leads towards "anticipation costs" while the other fork leads towards "failure costs." Anticipation costs are the money you spend in advance of a problem with the intent of avoiding that problem. Failure costs are the money you spend after a problem erupts and will largely be for repair but also for forensic investigation and the like.

These two costs are related, of course. Zero spend on anticipation will result in increased spend on failure (remediation) while near-infinite spend on anticipation will minimize failure costs. The optimal ratio between the two is dominated by the degree of collaboration you have with other information sources and sinks, that is, with the amount of electronic collaboration and coordination you have with your customers, suppliers, trading partners, and so forth. To get to our minimax solution – the maximum protection of information at the minimum cost – we need to place the two cost curves (for anticipation and for failure) on the same graph and take their sum. The minimum for that sum is our target; let's illustrate:

Figure 1: Anticipation vs. Failure



In the above, you can see that there is a bottoming of the cost curve where the costs of anticipation cross the cost of failure curve. That amount of information assurance is our minimax solution.

The most influential calibrator for the right amount of security spend is how much collaboration you have – which is to say the amount of information you have in play. For example, your organization is more "open" if you have a high number of electronic interactions with

[2] *"The Shrinking Perimeter: Making the Case for Data-Level Risk Management," Daniel E. Geer, Jr., ScD, Verdasys, Inc. January 2004.*
[3] *"Costs of Information Assurance," National Center for Manufacturing Sciences, August, 2002, p. 26., http://trust.ncms.org/pdf/CostInfoAssur-NCMS.pdf*

Zero spend on anticipation increases spend on remediation, while near-infinite spend on anticipation minimizes failure spend.

partners, customers, suppliers and the like. If you have very little collaboration, then the optimal spend for the sum of anticipation and failure will probably be small and made up largely of anticipation costs because with low collaboration your failure costs are low. At the other end of the spectrum, if your collaboration is massive then the sum of the two will indeed be larger and will be dominated by failure costs. The following three graphs illustrate this idea:

Figure 2: Low Collaboration

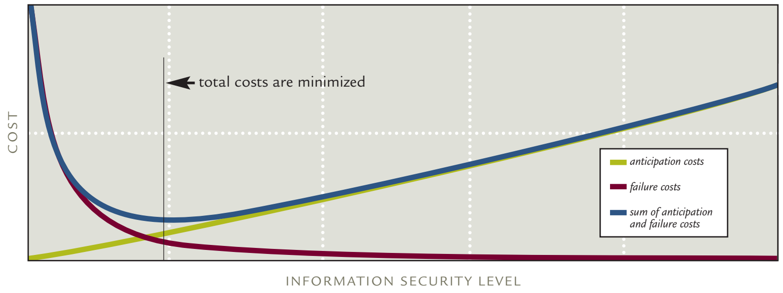


Figure 3: Medium Collaboration

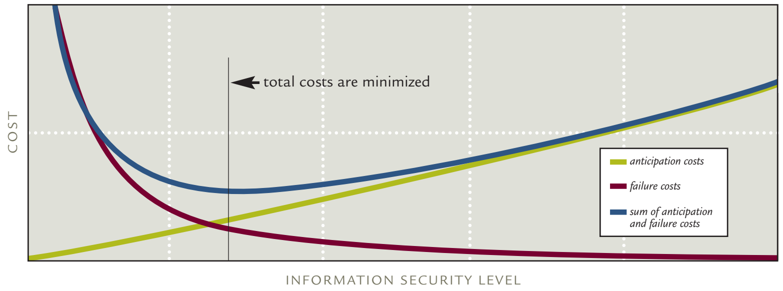
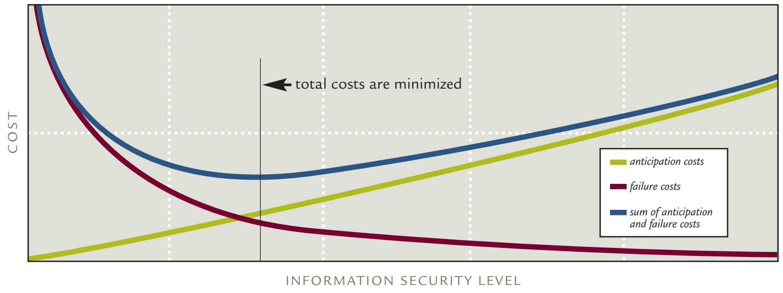


Figure 4: High Collaboration



As you can see, as collaboration increases the amount of information security that is optimal increases and so do the costs, but there remains a point at which there is a crossing of the cost curves and thus a minimax solution. We will explore both parts of this – the anticipation costs and the failure costs – below.

WHERE THE ANTICIPATION COSTS COME FROM

As a ratio, anticipation costs grow slightly faster than linear with the size of the enterprise. As a magnitude, they have been mounting upward in reaction to a steady stream of new risks. And they accumulate; when a new risk arises, someone will think up a new appliance or product to deal with that risk. If we want to do a good job, we need to understand the anticipation costs that are contributed by the most commonly installed mechanisms: antivirus, firewalls, intrusion detection systems, and patch management. For the purpose of this paper, we define:

Total cost of ownership (TCO) as the all encompassing set of costs — everything from capital to license fees to service costs to budgeted as well as unbudgeted labor costs.

TCO is the all encompassing set of costs — everything from capital to license fees to service costs to budgeted as well as unbudgeted labor costs.

We don't want to get into an argument about the methodology of computing the TCO; our discussion here needs to embrace the concept but does not need operational precision on how it is computed, subject to these observations:

- While it is difficult to precisely calculate the TCO of any technology, one can almost always establish a lower bound for that cost: "For technology X it has to be at least $Y."

  Example: The total cost of ownership of a PC rises as it grows older; the economically optimal time to upgrade is as soon as ongoing costs exceed switching costs.

- When product categories mature they become commodities. Once true, their purchase prices as commodities will be barely in excess of the cost of production and there will certainly be little difference between them, supplier to supplier. The direct implication is that any differences between TCO for commodities will be independent of purchase price.

  Example: Cellular telephones that are included in calling plans have zero marginal cost to the consumer hence the cost-effectiveness of any particular calling plan is independent of the purchase price of the phone.

- For any control system, exception handling is more expensive than routine management just as manual intervention is more expensive than automation. Under the continuing downward price pressure made possible through automation and all other things being equal, the spread between the cost of exception handling and the cost of automated routine management will therefore broaden over time, viz., the price difference grows wider. In TCO terms, the costs which deserve the greatest management focus are those which most contribute to that spread, which is to say the exception handling costs rather than the automated ones.

  Example: Credit card fraud costs card associations perhaps 75 basis points (of charge volume) while customer service is nearer to 200 basis points. The frontline defense against fraud is largely automated while customer service is almost entirely manual exception handling. Therefore, new cost avoidance, if rational, must be about customer service rather than fraud.

- In the security arena, the urgency of defense grows proportionally to the product of (1) the value of the data being protected multiplied by (2) the magnitude of the threat being arrayed against it. If either the value of the data is growing or the threat to the data is growing, then the urgency of defense will be growing. If both the value and the threat are growing, the urgency of defense-in-depth will be growing very quickly indeed. If an increase in value can be expected to draw more attacks (threat), then success in building value has an unavoidable side effect of increasing threat faster than that value.

Example: The more biometric data is collected and functionally available, the more it will be relied upon. The more it can be relied upon, the more it will be integrated into differing applications, becoming more valuable with each additional use. The more valuable these data become, the greater the motivation for an attacker to attempt improper access. The greater the number of applications through which an improper access can be made, the greater the overall risk. The overall risk is thus proportional both to the value, per se, and the number of entities that consider it valuable.

- All other things being equal, a minimal mix of products overlaps just enough that all requirements are covered. Such a minimal mix limits both capitalization costs and routine maintenance costs. If there were no exception handling we would be done. Where exception handling does exist, the minimal mix is also the optimal mix if it additionally provides that the exception handling of any one product is pre-empted by the routine processing of another. The optimal mix limits operational costs by limiting exception handling.

  Examples: A guard at a gate is effective. A guard at a gate plus a camera permits the guard to go to the toilet without having to dispatch another guard just to cover the time the first guard has the toilet door closed. If the camera malfunctions, the first guard can keep watching while the repairman works on the camera. Both the guard and the camera create exceptions but each covers the exception of the other.

The time has come to refocus information security on individual data objects and to do so at the point-of-use.

As of this writing, data represents both an increasing share of corporate assets and an increasing share of corporate risk. The time has thus come to refocus information security on individual data objects and to do so at the point point-of-use. In both policy and technical terms, increasing data-centered value and data-centered risk implies a defense-in-depth strategy, an array of protections in much the same way you might lock your car even though it is in a locked garage.

There are two prerequisites to our minimax solution: we must account for the cost of our defense-in-depth and we must establish a value on our information. Costing-out our defense-in-depth requires looking at the TCO for the existing melange of security products while paying particular attention to which of the component costs can be avoided, such as by better product mix and the refocusing on data. We will do that first. Valuing our information is more difficult, but it is tractable. We will do that second.

Comparands for Cost-Effectiveness

To state a conclusion at the outset: Popular security products are commodity priced, their TCO is dominated by the work factors of exception handling, and they do not counter-protect each other to any appreciable degree. Let's look at the TCO for those products, focusing on where the exception handling residuum is and what risks are created if ever one of those products were to fail.

To begin with, we have to have something to compare. Putting aside control mechanisms that no attacker would first attempt to subvert by brute force (like authentication), let's focus on the backstop mechanisms, the picket guards that nearly everyone has in place and with them compare the cost and the effectiveness of that cost. Because point-of-use is where the future lies, we start with the four main desktop mechanisms of protecting information: Antivirus (AV), Firewalls (FW), Intrusion Detection Systems (IDS), and Patch Management (PM).

Each of these technologies – AV, FW, IDS and PM – is less than a bargain. TCO across this group is substantially associated with the costs of management.

Antivirus

The problem statement for antivirus is easy to express and difficult to achieve, viz., to intervene in the process of software operation so as to recognize malware and to do so at minimum latency and minimum load. This is hard. It will get harder. At the limit, it is impossible. Quoting from the Information Research Council report on "Attacking Malicious Code"[4]

"In the end, determining if an arbitrary piece of code will behave maliciously in advance of executing it is as difficult as the Halting Problem. Thus, there will never be a complete solution."

This is bad news if the risk is growing, and of course it is. Some examples, perhaps: The interval between the appearance of a virus and the pervasive exposure of desktops to infection from it is shortening. Virus writers probably outnumber antivirus writers. Toolkits for virus writing are widely available. Virus payloads are growing more complex and harder to eradicate just as constant mutation of virus code ("polymorphism") is becoming more prevalent. Giving up on antivirus seems unthinkable, but planning for its failure is the essence of forward-looking security thought.

In the meantime, the antivirus market is commoditized and there are no margins to speak of. Subscription models are the only thing that works: a saturated market cannot yield substantial new sales and the need for prompt delivery of new signature rules virtually assures that only a push-model to existing subscribers will be able to keep up with the quickening pace of attacks. At the same time, cost per desktop cannot decline even if the risk is rising as margins are near zero as it is. The industry average for antivirus is $40 out-of-pocket per desktop year one;[5] TCO is dominated by the exception handling of fire drills and the compliance review of consistency desktop to desktop.

Firewalls

The problem statement for the firewall is parallel to that of antivirus: to intervene in the process of network packet delivery so as to recognize malware and to do so at minimum latency and minimum load. Quoting from the official Firewall FAQ.[6]

"What is a network firewall? A firewall is a system or group of systems that enforces an access control policy between two networks. The actual means by which this is accomplished varies widely, but in principle, the firewall can be thought of as a pair of mechanisms: one which exists to block traffic, and the other which exists to permit traffic. Some firewalls place a greater emphasis on blocking traffic, while others emphasize permitting traffic. Probably the most important thing to recognize about a firewall is that it implements an access control policy. If you don't have a good idea of what kind of access you want to allow or to deny, a firewall really won't help you. It's also important to recognize that the firewall's configuration, because it is a mechanism for enforcing policy, imposes its policy on everything behind it. Administrators for firewalls managing the connectivity for a large number of hosts therefore have a heavy responsibility."

The inspection needs to be made in near real time (as packets arrive at the desktop or network perimeter) so as to conclude whether or not to allow packets further forward progress. As there are many kinds of firewalls, it is hard to be more specific and yet be generally correct about firewalls as a class. At the end of the road, this, too, is impossible

[4] *http://www.cigital.com/irc/malicious_code.pdf*
[5] *Average of twelve samples, April 2004*
[6] *http://www.faqs.org/faqs/firewalls-faq/*

Exception handling costs deserve the greatest management focus.

(even for a stateful firewall) as an encrypted stream, once established, is by definition uninspectable. More generally, knowing that this or that content is bad is an even harder problem than the antivirus problem. As such, the firewall may be part of defense-in-depth, but it is not a free-standing solution. Protocols necessary for Web services require that the firewall be nearly transparent, e.g., SOAP.[7] Because those Web services protocols carry mobile code fragments, content inspection is more difficult than the antivirus problem.

As with antivirus, the firewall market is commoditized and margins are thin. Subscription models are not particularly meaningful and the already saturated market cannot yield substantial new sales. At the same time, cost per desktop cannot decline (even if the risk is rising) as margins are near zero as it is. The industry average is $40 out-of-pocket per desktop, year one.[8] As with AV, the TCO for firewalls will be dominated by change management though this will be less about fire drills and more about the cost of maintaining consistency and coherence whenever firewall count is high, as it always is in substantial enterprises.

Intrusion Detection Systems

The problem statement for an intrusion detection system is to gather and correlate sniffed evidence from the network to then be able to detect an attack while that attack is in progress. For the desktop, host-based IDS would be simultaneously looking at logs, data, network streams, and so forth. Some are signature-based in much the same way other desktop security technology is. Some are focused on anomaly detection since, as time goes on, the sophistication of the analysis can expand based on data accumulated to date. Anomaly detection, however, requires tuning by hand or a tolerance for false positives which will be otherwise generated whenever the user of the desktop does something unlike the day before. Operating at wire speeds is difficult, if possible at all.

Whether IDS is valuable or not is now a rather public debate with the Gartner Group leading the charge. In particular, Richard Stiennon, Research Vice President says:[9]

"Intrusion detection systems are a market failure, and vendors are now hyping intrusion prevention systems, which have also stalled. Functionality is moving into firewalls, which will perform deep packet inspection for content and malicious traffic blocking, as well as antivirus activities."

According to the Gartner Information Security Hype Cycle research, some of the problems associated with IDSs are:

- False positives and negatives
- An increased burden on the IS organization by requiring full-time monitoring (24 hours a day, seven days a week, 365 days a year)
- A taxing incident-response process
- An inability to monitor traffic at transmission rates greater than 600 megabits per second

Stiennon continues:

"Firewalls are the most-effective defense against cyberintruders on the network, and they are becoming increasingly better at blocking network-based attacks. To be considered as a challenger, visionary or leader, a vendor must have both network-level and application-level firewall capabilities in an integrated product. Vendors that have only one or the other will be niche players."

[7] *http://www.computerworld.com/softwaretopics/software/appdev/story/0,10801,49374,00.html*
[8] *Average of eight samples, April 2004*
[9] *http://www.gartner.com/5_about/press_releases/pr11june2003c.jsp*

Information that is protected is information that has value.

While that is harsh, it does illustrate a substantial overlap in the functionality of a firewall and an intrusion detection system. It is hard to see where progress is going to come from and for the same reasons as before: the margins are thin and the penetration is moderately high. Subscription models are not meaningful at all and the market won't yield substantial new sales. At the same time, cost per desktop cannot decline (even if the risk is rising) as margins are near zero as it is. The industry average is $32 out-of-pocket per desktop, year one.[10] The TCO is thus dominated, as it is with firewalls, by change management rather than fire drills but with the cost of investigating false positives added in. If anomaly detection is to proceed at the individual level, i.e., if tuning for anomaly detection is to include what is normal for the individual at his or her own desktop, the TCO will also include keeping such tuning up to date and, perhaps, making sense of the differences in "normal" desktop-to-desktop. Note that per-person anomaly detection will not be possible in a serial re-use environment ("free-seating").

Patch Management

Patch management is a different beast, but it, too, is aimed at the desktop though it necessarily will include a server component as well. The problem statement is to ensure that the files that are on the desktop are the right files within the particular interest or focus on files that represent security or functionality risk. Because patch management is not really management without a measure of control, much of the problem is to be sure what files are on the desktop, to compare it to what is thought to be the right set, and to move copies of what should be there but is not. (We put aside the question of whether it is possible to operate a patch management system that is not itself a latent source of vulnerabilities to the network and hosts on which it runs, such as if the patch management system is itself penetrated by an attacker.)

The PM work factor will be dominated by peak-loads as the patches that are the most important to put out are the ones intended to cure security holes, per se. This may include time-of-day controls as well as latency minimization which will interact whenever the patch in question approaches a threshold of severity such that the attack risk of not patching exceeds the interruption risk of patching delay-free (see below). As such, keeping a close eye on what files are present on the host cannot rely on memory of what was last there, especially now that Microsoft and others have committed to "patch rollback" as a feature of all distributed patches.[11] Instead, PM must rely on inventory and replacement of files at a distance.

The current pricing of patch management systems is low not so much as it is due to commoditization pressure but because of the slight functionality that is really involved. The industry average is $25 out-of-pocket per desktop, year one.[12] As with AV, exception handling will dominate the TCO as the urgency of installing patches can be quite high indeed. For an enterprise of even modest size, there is never a moment when all desktops are at the same version, fully reachable by network systems administration, and ready to receive repairs that may require termination of currently running programs. Patch management, in other words, may be the most dominated by exception handling of them all as it is, in and of itself, a vehicle for handling exceptions even to its problem statement itself.

Put differently, just doing a patch to a large enterprise requires at least $10/desktop in direct costs[13] to do and that is with automation while, much more importantly, the indirect

[10] *Average of six samples, April 2004*

[11] *http://www.microsoft.com/presspass/press/2003/oct03/10-09SecurityInvestmentspr.asp*

[12] *Average of twelve samples, April 2004*

[13] *Median lower-bound estimate drawn from private communications with six representative organizations of substantial size.*

costs include such hard-to-measure components as the downtime that installation requires and the pre-deployment testing costs. For a "red" patch, there may not even be testing which is simply a cold-hearted risk trade-off question: Is the risk of breaking things with a patch we have not tested greater or less than the risk of attack due to the absence of that patch? Of course, decisions like that will interact with questions such as, "Who's in charge here?"

You Get What You Pay For, But Not Forever

It is time for an assertion: Each of the four main information security technologies is less than a bargain; their pricing doesn't leave much margin for R&D, there is an externally imposed, continual creep in their frank capacity (like firewalls) to add value, and what's more, the requisite time-latency of each (to do its job) is getting more challenging by the day. The up-front, out-of-pocket costs are virtually irrelevant to any calculation of TCO as the substantial costs of management far overshadow the purchase price. The customers and suppliers of these four alike undertake to keep these mechanisms running in the face of perpetually mutating threats with ever more challenging time scales. The curve of real TCO is, over time, rising. Its rate of rise will grow steeper if the cost of R&D must be amortized in. Because the cost of anything is the foregone alternative, the opportunity cost to in-customer staff to manage an ever harder problem would, if recognized in TCO, also steepen the curve. When a TCO curve bends upward over time, either the benefit must bend faster or the enterprise will accumulate diseconomy. While there is no doubt that security technologies have long struggled to be anything other than a diseconomy, that is no excuse not to seek ways to make the curve of costs bend away downward when it is prudent to do so. If this is not done, then security eventually becomes an enemy of productivity, which is to say wealth creation, which is untenable over the medium term much less the long. Something has to give, the only question is when.

AV, IDS, and FW all assume that the outside world has to be kept out. PM is a rear guard soldier that just takes orders and cleans the latrine. Is there a unifying abstraction that could make this problem tractable and, could displace technologies before they lose big? The answer is "Yes."

From a security point of view, a computer is basically three components: communications (bandwidth), storage, and CPU. The only one of those three that has economic value is the storage in that it holds information which is the economically valuable part while the rest just service that value. AV, IDS, FW and PM are all about protecting the economically valuable part – all about protecting what is in the storage. Frankly, information that has security is information that has value; information that has no security has no value. Sometimes what is in that storage is a program (an executable), sometimes it is a trade secret (information you own), and sometimes it is a medical record (information entrusted to you), but AV, IDS, FW and PM are each and severally about protecting that storage. If we could make protecting the storage a unitary function – one that both made mucking with it hard and getting away with it harder – we'd have something.

AV, IDS and FW all try to keep files from being introduced that shouldn't be introduced, or modified that shouldn't be modified. If they fail, and their task is getting harder all the time, then one of two things happens: A file gets modified or a file gets transmitted to someone who shouldn't have it. If you can just prevent files from being modified that shouldn't be modified, or transmitted that shouldn't be transmitted, and do it directly then you have something. You could think of it just in terms of backstopping the inevitable failure of AV or IDS or FW, or you could think of it as replacing some or all of them.

If you can just prevent files from being modified that shouldn't be modified, or transmitted that shouldn't be transmitted, and do it directly then you have something.

PM tries to inventory the files on a system and compare it to a known gold standard. Why stop there? Why not inventory all the files and simply always know what is present? Then if you actually do want to install a patch, then you need only ask who has the patch and who doesn't – something your file inventory knows already. If your file inventory keeps itself up-to-date, the work list for patching is always present but it is no different from the work list for upgrades, fresh installs, or repair. It is one function based on knowledge, which is to say power.

Each of those systems – AV, FW, IDS, PM – does an incomplete job and what news it brings you is bad news, while with the small exception of PM, doesn't even help you fix anything.

Sum up the first year out-of-pocket costs of doing all four and it simply doesn't matter:

Antivirus	*$39.63*	*average of twelve examples*
Personal Firewalls	*$39.83*	*average of eight examples*
Patch Management	*$24.53*	*average of six examples*
Intrusion Detection Systems	*$32.02*	*average of twelve examples*
	$136.01	*out of pocket*

But add in personnel at 1 FTE (at a real cost of at least $100,000/year) for every 200 desktops and you have a recurring $500 per desktop per year just for direct costs of security. Multiply that by two to get a minimum bound on the additional impact of intangible costs like testing, downtime, and the spare capacity needed for fire drills and the TCO for security is a substantial and increasing fraction of the total TCO for the desktop. (An academic environment, like U of Washington estimates $3K/year while Fortune and The Economist estimate $4-8K/year.)[14]

In all that (AV/FW/IDS/PM), none of these products provides meaningful data for management beyond just counting the number of alarms, or the number of patches to be pushed down, or a list of potential intrusions that have to then be investigated with scant data to work from. In other words, each of those systems does an incomplete job and what news it brings you is bad news while, with the small exception of PM, it doesn't even help you fix anything.

WHERE THE FAILURE COSTS COME FROM

When security fails, viz., when AV, FW, PM or IDS fail, what happens? At its most basic, one of two things happens: either information is destroyed or information is sent to places it shouldn't go. That is your cue to ask, "So what is that information worth?"

We need measures on what information is worth or, as we alluded to at the outset, at least lower bounds on what information is worth. There won't soon be a precise answer the way there is to the price of the morning gold fix in London; in some cases, the wrong bit of information being sent elsewhere is so catastrophic that whatever equation we are working on is more than overwhelmed by the one single cost. In others, the cost of damaged information might be mere drudgery for someone to re-load from other media. In yet others, the exposure of information may create culpabilities quite outsized when compared to the real damage. One thing, though, is for sure: Evidentiary-grade forensic data will be the basis for avoiding a repetition and forensic data is three orders of magnitude more expensive to recover than to capture.

[14] *http://www.washington.edu/nebula/tco.html*

Information Value (positive)

In literature, and we mean the academic literature as well as the policy literature (including government), all evaluators agree that information is valuable. Sadly, that is as far as it goes, i.e., they all come to idiosyncratic conclusions some of which smell like justification for budget. This is not a crime, but it makes generalization from them tough.

Information losses – including diversions – can affect reputation, your ability just to operate, and your future, the last especially as it involves intellectual property.

In the realm of statistics, it is often the case that a difficulty in figuring the value of some variable "X" can be avoided by figuring the value of "1-X" which is to say that solving a problem may involve changing your point of view to another one where the computation is tractable and the solution equivalent.

For example, "How much is your brand worth?" doesn't get a ready consensus answer in any board room. On the other hand, asking, "Knowing what you know now and starting from scratch today, how much money and time would it take to build a brand as good as the one you have now?", is often the change of viewpoint that gets you, if not a consensus estimate, at least a consensus lower bound. In other words, "We don't know what our brand is worth" becomes "It is worth at least $60M." When that lower bound is sufficiently large to make whatever decision is on the table, you are both using reputation value as a numerical estimate and not getting hung up on precision. Precision is good, decisions are better.

So, back to the question of the value of information. Let's assume we cannot just measure it with a yardstick. What can we do? We can ask what costs we would incur if we did not have the information (or if someone else did). We can ask what replacement information of equally trustable guidance would cost us, starting from where we are and knowing what we now know. That we can probably do, and that is what we suggest that every board in the country do or delegate getting done. This is the idea whose time has come.

Information Value (negative)

When considering the inverse value of information as we are suggesting here, it may help to break information up into categories. It is, after all, easier to chew bite-sized chunks. Information losses (including diversions) can affect reputation, your ability just to operate, and your future, the last especially as it involves intellectual property. So think of it this way:

Reputation Value What would it cost to build a reputation as good as the one you have now? What information would, if lost or diverted, cause you to have to rebuild your reputation? Is that information surrounded by a wall and, if so, how high? Is the wall one that is walling in (against internal attacks) or walling out (against external attacks)? Would you promptly know if your information was gone, or is it more subtle and the injury might go unnoticed for long enough that the costs of reputation repair were themselves proportional to how long that period before detection had persisted? Does the value of your reputation already exist on your balance sheet and, if so, is that "good will" or another form? Does your reputation depend on public trust and what information would the public care about you losing? Does your regulatory environment assign liability, per se? Do you know how much information there is? Do you traffic in the information on which your reputation is based, such as research reports drawn from pools of raw information that you develop?

Operation Value What information do you use daily to run your machines, to answer your phones, to deploy your staff, to file your taxes? What information is operationally essential? What information is sufficient to get the job done and what information is necessary but not sufficient? If you had to rebuild that information, would it just be an historical treasure hunt through miles of backup tapes or does it exist in more complex forms like the coordinated settings of the many details that make just-in-time delivery possible? If just a little bit of the information is poisoned, would you notice and would your processes self-correct? If so, how much information poisoning is needed before you lose the ability to self-correct and must instead re-create? Would you know that poisoning had occurred quickly, or would your products be in the hands of your customers before you realized that the wrong resin went into the wrong batch of thermoplastic? Do you operate under requirements that demand continuity of measurement, day by day and hour by hour? If you lose that continuity, how much burn-in time do you need to restart – an hour, a day, a month, too long to think about? Is your operational information naturally a trade good with your suppliers or customers and what happens if they find you've lost theirs or sent them poisons?

Evidentiary-grade forensics data will be the basis for avoiding a repetition and forensic data is three orders of magnitude more expensive to recover than to capture.

Future Do you have trade secrets or patentable inventions in progress or laboratory results in progress? Is your field competitive enough that losing the kind of information that deserves the name "intellectual property" would itself be an irreversible catastrophe? Is that information required to be dispersed and does that dispersion include multiple departments or multiple jurisdictions? Is your market position based on your market lead in the lab? If you had to recreate as much market lead as you now have from scratch could you do it and, if so, what would it cost? Is there a first mover advantage in your field, especially one where standardization tends to be a de facto benefit of being that first mover? Can a diversion of your information plausibly enable someone else to overtake you and claim, instead of you, that first mover advantage, that market-based standardization? Is the core value that your stockholders and investors see in you represented by your information? Do you value your intellectual property consistent with the market's valuation of your company? If less, who holds the responsibility to explain should explanation be required? If more, why do you value your intellectual property more than the market does (which is perhaps the ultimate insider question)?

It is of course entirely possible for a fool to ask questions that a wise man cannot answer, and the above is a lot of questions even if it is nowhere near complete. Our point though is that a straightforward "So, what's your information worth?" isn't likely to help – it is early and no one yet knows how to do this in a consensus fashion. But a sense of what does information enable by looking at the inverse of what the absence of that information disables can and will lead to insights on the appropriate level of protection and it can do so today.

We suggest that almost no one is adequately protected, but almost no one is going to be adequately protected when they haven't gone through an exercise like the one we are suggesting. Then, and only then, can we precisely look at the TCO for protection models. When we do, we'll find that models that value information the highest will, simply by logic and necessity, value protections that are about data at the point-of-use, protections which do not actually make any distinction between insiders and outsiders, and protections which defer exception handling to times of genuine necessity as determined at home, not by the failures of suppliers and counterparties.

A UNIFYING ABSTRACTION

We have talked about anticipation costs and did it first because AV, FW, IDS and PM represent what everyone is already doing. We have talked about failure costs in order to assess the value of information and the corresponding need to protect it. We need now to talk about a

Agent-based capture of file usage at the point-of-use plus risk appropriate triggers... supplant the need for AV, FW, IDS and PM on TCO grounds.

way to pull these two together, to find the minimax point where the sum of the anticipation costs and the failure costs bottoms out and the ratio of security achieved to cost expended is maximum. We need to find our economically minimax solution.

Recall that the more open the corporation's business model the more information security contracts to the point-of-use.

As…	**Thus…**
…at the point of use, the granularity of information is generally a file,	*… security arranged around the individual file is security at the most appropriate level of abstraction.*
…security that gets in the way is security that is circumvented,	*… data collection must be no-load and inescapable while data analysis must be off-platform at a management console.*
…access control neither scales nor reaches files already on the desktop,	*… point-of-use accountability must be the model for our design.*
…reconstruction of event traces is a thousand times more expensive than live capture and retention,[15]	*… accountability must do full data capture at the point of use.*
…the more sophisticated the attack the more it is undetectable,	*… design must interdict data operations at their moment of use, viz., point-of-use in both space and time.*
…nothing is perfect,	*… the security system must signal its own failures.*

Of course such a solution exists; we would not be writing this if it didn't. It is called "Digital Guardian" or DG for short. Here is the claim:

Agent-based capture of file usage at the point-of-use plus risk appropriate triggers that respond in context to user, application and intended action on data supplant some and in time all of the need for AV, for FW, for IDS, for PM, and it does so on TCO grounds.

The comprehensive collection of activity data at the point-of-use provides for enterprise wide audit to be used in:

1 Understanding how information is misused to ensure that security policy is based on knowledge rather than assumptions of where risks truly exist.

2 Ensuring that the process and policy dictated by regulatory and internal requirements are in fact followed by those entrusted with access to information.

3 Because no security is perfect, when business needs demand that risky activity be permitted and technology failure or human error result in an incident, that incident is immediately recognized and can be traced to its source for timely mitigation.

If the idea of collecting this much activity data using traditional logging technologies makes such an approach seem unfathomable in scale, fear not. By correlating low level OS and application actions into the higher level user actions they represent and ignoring OS and application activity of little interest through simple filtering techniques, a data reduction

[15] *This is an underestimate; competent forensic specialists will cost in excess of $2,000/day, perhaps much in excess. Reconstructing one day's worth of activity for an active machine where misuse has occurred will require more than one day of effort. Amortizing the cost of Digital Guardian over just the first year of operation would be substantially below $1/day for full data capture and analysis. Hence a minimum of three orders of magnitude. In personal communication one active practitioner recommended four, not three. At either figure, the point stands.*

of many orders of magnitude is achieved at the point of capture. In practice, an average user's daily activities can be represented by the equivalent of a few Web pages worth of data. This makes the transmission of such data to centralized storage over a 12-24 hour period negligible in terms of network load and the issue of storage requirements a question of the length of time live data is kept online. Since it's the context of a user's actions that we are tracking and not the content and most of the time is spent ignoring activity of little interest, local resource utilization (CPU, I/O, etc.) is also negligible and orders of magnitude below existing technologies.

Trust, but verify.

The ability to compose and enforce risk appropriate rules and policies at the same point-of-use provides for an approach to securing information that encompasses not only the network, but every other subsystem involved in the movement and/or removal of information. By risk-appropriate, we mean the choices of:

1 Silently alerting security staff when surveillance is desired without user involvement.

2 Prompting a user before actions are permitted when education and awareness are desired.

3 Blocking user actions when the activity being performed is too risky to allow and has no business justification.

The visibility offered by a policy engine at the point-of-use enables a level of granularity in composing policy that can make decisions in context using the user, the application, the transmission method (network, removable media, clipboard, printing, etc.) and the piece of data involved and can't be easily blinded by encryption as it watches before encryption at the network layer occurs.

Digital Guardian offers considerable information protection benefits directly tied to low TCO:

- The desktop price is approximately $100/seat and moves downward based on volume.
- The computational footprint is 1/10th of AV alone much less the sum of all four.
- The data captured is universally valuable for far more than information security though there is more than sufficient justification in the security aspect alone to make this cost effective.
- DG has zero emergency updates, but it does prevent whatever failure AV or FW or IDS would otherwise let happen.
- DG collects forensic data as a matter of course and at near zero operational cost.
- DG defers processing of forensic data to times of necessity.

To put the hardest edge on this whole question, the one sentence that makes plain the real, lasting and volatile tradeoff the security industry has to make (given all the whirlwind of technical details that we are leaving out because for this question they just don't matter), it goes like this: Our (everyone's) remaining choice is between surveilling data or surveilling people; DG surveils data. More straightforwardly, if you watch the data you don't have to watch the people.

This is the future. It is cost-effective. Its security is better. It makes the hard tradeoffs in the open. It is built for a world where information is valuable and in motion. It is "Trust, but verify."

APPENDIX — POINT PRODUCT COMPARISONS

Comparison of DG against AV

AV exists to block malware. Other than saturation ("resource starving") attacks against the local CPU, AV exists to prevent the insertion, deletion, or access to file-level resources. Because AV must check amongst an ever-increasing list of signatures (circa 5,000 as of this writing), its work factor is controlled by the attack community. Where AV has to do genomic analysis of the code, DG has only to do genealogy. In an environment where going directly to whitelists is desirable, the need for AV evaporates completely. In an environment where code can do whatever it wants to do except for what is on the blacklist, DG supplants AV as it offers the same function, more effectively and for 10% of the CPU footprint. Where constant updates of what to look for consumes personnel resources and introduces regular emergencies, DG requires no such intervention – ever more inventive ways to do things that are already blocked creates no emergencies. Where sophisticated environments run AV on outbound communications just like inbound, they do it to protect against failure of their inbound AV.

Comparison of DG against FW

FW exists to do content inspection of inbound packets and to simulate, via statefulness, what is going on in this or that application without being privy to the business logic in those applications. Assuming that no one would do this if they had any faith in their applications whatsoever, the value of the FW is inversely proportional to application trust. So much for cost-effectiveness in the big picture. Because a FW will always be underfoot, it will necessarily either miss things or, if behaviorally restrictive, interfere with novel but legitimate activities unless trained. Like AV, frequent updates are the rule and the performance of these carries with it a penalty that is in proportion to the virulence of the present disease whatever that may be. DG, by contrast, does not attempt to save applications from themselves but rather to save data from the applications. DG doesn't trust them either, but it doesn't get into a packet-level debate over the question, it just protects the files. This is peace through strength versus Rent-a-Cop.

Comparison of DG against IDS

Given that the FW sector is already being cheered for encroaching on the IDS sector to the degree that it is, one might ask why it is that we now beat a dead horse. The answer is that the surest sign of an intrusion is on a small list of evidences each and every one of which involves a file touch in some way. Back doors installed? Not if file changes and insertions are surveilled and blocked (the former for your honeypot, the latter for your production environment). Buffer overflows leading to supervisory access on the public network rather than the management net? Not if access paths are surveilled in real time. Misuse of the machine by its legitimate user? IDS? No answer, no help. But DG, because it operates at the individual data object level, blocks the effect of an intrusion. If you believe that 80% of the problem is insiders, not intruders, then by solving the 80% you largely blunt the remaining 20%, which is surely cost-effective.

Comparison of DG against PM

PM does little beyond inserting files into a file system on the basis of assessed need. The need assessment is an inventory that DG maintains as a matter of course, so PM doesn't add any functionality there. File insertion is already built into any environment that has the ability to do remote system maintenance, so PM is really just a mix of file inventory and file insertion, both of which are available without it. So what is its value add? Hard to see it unless the management console is an overwhelmingly nice piece of work. DG has a very nice management console that can not only tell you who needs this or that patch file but even before the patch file is available it can tell you who does or does not have the vulnerable version of the soon-to-be-patched file. A sorting pass over the file inventory tells you who needs the patch. A different sorting pass over the file inventory changes tells you who's been naughty versus nice in taking patches. Yet another sorting pass over the file inventory tells you the effectiveness of your patching discipline. All of this is a side effect of better security at the point-of-use, which is cost-effectiveness personified.

ABOUT VERDASYS

Verdasys™ information security solutions enable the higher standard of due care that today's business environments demand. Advanced point-of-use data monitoring and control solutions – patents pending – enable management to confidently prevent threats to information, inside or out, audit information use and access for compliance with regulatory requirements, and quickly and easily trace incidents back to their source. Verdasys Digital Guardian™ is used today by:

- Leading health insurers to protect the security of proprietary software applications while employing offshore contractors
- Global media/entertainment companies to prevent intellectual property loss, and
- Large manufacturers to ensure consistent and proper usage and containment of regulated and proprietary data.

For more information, please visit www.verdasys.com, or call us at 781-788-8180.

Verdasys, Inc.
950 Winter Street
Suite 2600
Waltham, MA 02451
781-788-8180 Tel
781-788-8188 Fax
info@verdasys.com
www.verdasys.com