

WHITE PAPER

Securing the Point of Use

The New Foundation for Data Security

Daniel E. Geer, Jr., ScD
Vice President, Chief Scientist
Verdasys, Inc.
May 2005

VERDASYS[™]

Data Security at the Point of Use

Information security is what ultimately distinguishes information that has economic value from information that does not.

Introduction

Information is a growing fraction of corporate wealth. Quality information is hard to acquire and easy to lose. The good news is that information is easy to move and easy to alter – but in a world of burgeoning threats to data, that is also bad news.

As complex as information security needs are, more complexity is not the answer. All of us on the good side need help that isn't infantile, rocket science or pandering.

This paper is that help. It is a way to make sense of a messy reality and thus make action un-messy to take. It explains why to strategically focus security on data at the point of use, how to economically optimize the ratio of security-spend to security-value, and where to tactically deploy your forces. It is a complete picture.

Information Security Matters

Information security is an economic issue, just as quality and reliability are economic issues. Every company has information that in and of itself is a corporate asset. The loss of a corporate asset has a negative impact on the corporate balance sheet, whether we “realize” that loss or not.

While the means to achieve security or quality or reliability may be technical, the goals are economic. In other words, information security is what ultimately distinguishes information that has economic value from information that does not.

Information is only of value when it is used. When it is used, we can say that it is “in motion,” whether we are reading a file, sending a document by email, using cut and paste to assemble a body of work, running a program, converting bits to paper, or any myriad of tasks. The point at which information changes state from “at rest” to “in motion” is the one place where all activity can be both seen and judged against policy and risk. It is here – the point of use – where our security efforts must now be concentrated, the point of use.

Data has value and data is mobile. That statement is almost redundant, because data that is never mobile is data that has no value while, data that has no intrinsic value is unlikely to ever be used. These two statements – that data has value and that data is mobile – are thus as tightly conjoined as the two faces of a coin.

If that is true today, it will be truer tomorrow. For some time now, industrial research laboratories have been making raw computing horsepower cheaper month over month. More importantly, they have been making storage capacity cheaper at an even faster clip and bandwidth even more so. As rules of thumb, computing per dollar doubles every 18 months (Moore's Law), but for storage it doubles every 12 months and for bandwidth, every nine. This means the optimal computing plant becomes significantly more data-rich over time and that data, despite growing in volume, becomes ever more mobile. In today's information-focused business environment, this is brute economics, and it must be recognized.

So, if data has value and if data is mobile, then all remaining questions echo the earlier claim, namely, that information security is what distinguishes information that has economic value from information that does not. If you cannot control it,

how can you say that it is yours? How do you guarantee that it will be used to your benefit? How do you know your web of risk and responsibility in owning that data is adequately hedged? How can you even know where it is at all times? You can't.

Add to this environment the fact that the threat to data is rising, and you have a greater challenge than ever before. Attack tools encapsulating one-of-a-kind skills are widely available, and software patches are being reverse-engineered in record time to facilitate exploits. As companies increase collaboration with partners, suppliers and other third parties, the operational distinction between the inside and the outside of the corporate perimeter becomes ever more difficult to define – and makes security more difficult to enforce.

It's a fact of nature: When threat goes up, the defensible perimeter contracts. Armies know this. Wildebeeste know this. And those responsible for information security know this. They, too, shrink their protection perimeter when threat is rising.

But shrink it to what? You can probably guess. The curve of threat-driven perimeter shrinkage crosses the curve of increasing information value at a point where data value and mobility can mean only one thing: The perimeter shrinks to individual data objects. The mechanism for enforcing that perimeter moves to the data's point of use, the place where data has a state change from quiescent to volatile. The simple, cost-effective approach is to move protection to where the data is – rather than moving the data to where the protection would be convenient.

Thinking the Problem Through to the End

Take as our design problem the protection of individual, file-level data objects at the point of use. The general solution to security design problems has always had two parts: (1) Trust the people you have to trust, but (2) Make sure that they are who they say they are. In practice, this is one part "authentication" (*Who are you?*) and one part "authorization" (*Given who you are, what can you do?*). Collectively, this is known as access control.

However, in a world of high value, high mobility data, access control does not provide enough security because it ignores the potential threat from insiders. At the same time, it fails to provide cost-effective security because it does not scale.

Consider the access control model. Picture a matrix: One row for each person who can request a system resource; one column for each system resource. The number of boxes in this matrix is the product of the number of people times the number of resources. If you double the size of the company, you quadruple the number of boxes. In other words, if there is a minimum managerial cost to maintaining a check in each box, the cost of maintaining the matrix grows faster than linear with company growth. Any cost that scales faster than linear is a barrier to said growth. When security is that barrier, people simply work around it.

A similar argument applies if you are busy making your company more secure by subdividing those rows and columns into finer grained access control. Even without the issue of corporate growth, pushing access control too far ensures that the result is uneconomic. The only question is when it becomes so.

When threat goes up, the defensible perimeter contracts. Armies know this. Wildebeeste know this. And those responsible for information security know this. They, too, shrink their protection perimeter when threat is rising.

Accountability steps in where access control leaves off. In a free society, there is huge efficiency in not having to ask permission for every little thing – but if and only if there is a high probability that when you misuse your freedom, you lose your freedom. This is the definition of accountability. In a business environment, as we contract our perimeters to individual data objects, we cannot afford a matrix of checkboxes for every object, so we have to rely on accountability rather than access control.

You Can Observe a Lot Just By Watching

In the 1990s, the commercial sector caught up with the military sector in cryptography. Crypto is now everywhere – cheap and unremarkable. In this decade, the same thing will happen to what military folks call “traffic analysis” – the act of paying attention to what you can see and measure. If we are going to contract our protection to individual data objects and invest in accountability rather than more access control, then we can only begin by tracking all handling of data.

How can you do that? It’s simple – and if it weren’t simple it wouldn’t work. Complexity is the enemy of security, and a complex security product risks being an oxymoron.

The answer is to intercept all transactions that involve files. Think of it as event detection. The event records are filtered and correlated at the time of capture to distinguish standard OS and application activity from user-initiated data use for a drastic reduction in volume, making such capture feasible at enterprise scale. The audit trail is then periodically compressed, made tamperproof and off-loaded to a central archive.

Because this capture occurs in real-time, you can also react to events at any number of levels, making the audit trail actionable through rules and policies. The reaction to an event should be risk-appropriate and tuned to business needs. This can range from passively recording the event and issuing a silent alarm, to promoting user awareness through warnings that ensure informed decisions are being made, to outright blocking when required.

The point is: You need to get the event log. If you don’t get it while it’s fresh, you won’t get it at all – or you’ll have to pay a thousand times as much to reconstruct it as it would have cost to capture it in the first place. And you need to match policy enforcement to business needs, so risk can be managed without impeding productivity.

If You Can’t Win, Change the Rules

In all settings where engineering meets policy, getting the problem statement right is job one. If the problem statement is a direct characterization of an actual pain point, then a cost-effective solution must necessarily be narrow.

This explains why there are a large number of security products built around a small number of mechanisms. There are too many products to list, but the preponderance of them address on-the-wire content inspection, statistical anomaly detection, and/or signature finding. Let’s examine these products one-by-one.

Complexity is the enemy of security, and a complex security product risks being an oxymoron.

*You need to get the event log.
If you don't get it while it's
fresh, you won't get it at all –
or you'll have to pay a thousand
times as much to reconstruct
it as it would have cost to
capture it in the first place.*

Content inspection is frail and unscaleable. Though you must ignore USB tokens, CD-ROM burners and even floppy disks, it remains seductively attractive that some sort of network appliance could listen to all the packets on the wire and selectively interdict the ones containing corporate intellectual property or naughty bits. With network traffic rates rising faster than CPU horsepower, wire-speed content inspection cannot scale cost-effectively. Encryption defeats content inspection even when that cryptography is no more robust than pig Latin. Message fragmentation defeats content inspection even if the fragmentation is provided by network routing and not an insider or intruder who carries only a little out each day. Finally, content that has no regular and predictable form, such as digitized media, is not only hard to characterize sufficiently to motivate inspection but is the canonic vehicle for steganographic¹ concealment of one form of content within another. Content inspection may be an adequate tool of choice when the opponent is ignorance or inadvertence – but not when your opponent is neither ignorant nor careless. Content inspection is thus inadvisable unless just looking diligent is good enough.

Statistical anomaly detection (SAD) is always interesting because every body of data on the planet can generate interesting hypotheses. That is what SAD does – it generates hypotheses. Unfortunately, it is up to you to purge SAD's spew of false positives. Just as "New!" doesn't necessarily mean better, an anomaly is nothing more than an invitation to expend investigative energy without knowing why. When you are trying to find a signal amongst much noise – for instance, if you are searching for signs of extraterrestrial intelligence (SETI) – that's as productive as you can get, but when you are trying to protect against effects that you understand caused by mechanisms that you do not, it's an effective way to proceed.

Worse, anomaly detection followed by anomaly analysis is a success when it turns out that nothing bad is happening. This tends to dampen the enthusiasm of analysts, not to mention capping their number in the budgetary sense. If you are protecting the nuclear launch codes, it's all well and good. If you are trying to make money in the real world, it's simply a way to waste time.

Signature finding is certainly a venerable approach, and when the number of signatures to find is reasonable, not to mention bounded, it does give one a certain independence vis-à-vis methods of delivery. But there's the rub – the number of signatures today is neither reasonable nor bounded. Symantec estimates one new Windows virus every three hours, and every new application presents new opportunities for virus writers. Signature detection presupposes that the signatures are universally known before the attack occurs. There is little doubt that the fraction of total attacks due to unknown attack vectors is rising, just as there is little doubt that unknown unknowns are as important in this space as they are in matters of national defense.² In short, what was once a good idea for a world that we no longer inhabit is now a fading light in the world we do.

An effective security strategy must protect data at the point of use. While it must be an observer, more protection value is obtained if it is also an enforcer of policy. If it doesn't observe, then it cannot control.

If, however, the problem statement is goal-directed at information protection rather than the *maladie de la saison*³, then a cost-effective solution will be adaptable above all else, because threat evolves. Hence, the fundamental claim around Verdasys' Digital Guardian security platform: Where future-proof information protection is the goal state, a reference monitor strategy is the only real choice. It alone gets the abstraction right, and it is available today.

A fool can ask questions that a wise man cannot answer. “*What's your information worth?*” is one of those questions, but looking at what an absence of information would disable should indicate the appropriate level of protection. If you take a look at the TCO for potential protection models, you will find that models that value information the highest, inevitably coalesce into protection of data at the point of use. This type of protection requires no distinction between insiders and outsiders and minimizes exception handling.

Any security budget is the sum of anticipation costs (preventing trouble) and failure costs (cleaning up trouble). The key is to minimize the sum of the two, while keeping in mind that they are negatively correlated: Infinite spend on prevention means never having to clean up, while zero spend on prevention assures significant, imminent cleanup costs. Let's now discuss a security system design that pulls these two costs together and finds that “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.

A Unifying Abstraction, in the Flesh

A computer is composed of a CPU, a storage device, and some network bandwidth. All of the economic value is in the storage; everything else is just residual resale value. Economically speaking, a security strategy begins and ends with data protection. Securely speaking, an effective security strategy must protect data at the point of use. While it must be an observer, more protection value is obtained if it is also an enforcer of policy. If it doesn't observe, then it cannot control.

As access control neither scales nor reaches files already on the desktop, accountability is the core of our design. As security that gets in the way is security that is circumvented, data collection and data analysis must be split between agent and console respectively. As reconstruction is a thousand times more expensive than retention, full data capture is a requirement. As more sophisticated attacks are less detectable, our design must interdict data operations at their moment of use. As nothing is perfect, our security system must signal its own failures. As the point of maximum risk is at the point of use, our security must be host-based.

A host-based security platform performs three functions: monitoring, reporting and control. The use of data is relatively easy to monitor, but if and only if there is an agent installed on the host – one that is close enough to the iron to miss nothing that happens to the data. This is – or was once – known as a “reference monitor,” and it is an idea whose time has come and gone and come again (though with the small amendment that its original purpose as purely access control must now be

subtly reconfigured to support accountability). To deserve the title “reference monitor,” a proposed technology “...should be: (a) complete (i.e., it mediates every access), (b) isolated (i.e., it cannot be modified by other system entities), and (c) verifiable (i.e., small enough to be subjected to analysis and tests to ensure that it is correct).”⁴

What has changed to make this idea essential and practical in today’s world? First, rising threat has, for better or worse, balanced the cost of implementation with the cost of failing to implement to the extent that implementation is now the preferred economic alternative. Second, Moore’s Law now delivers enough horsepower to permit reference monitors to have no noticeable performance impact.

The notion of a reference monitor – a distributed, host-based data security platform focused on accountability – and it has arrived in the form of Digital Guardian. Digital Guardian is a recording reference monitor: Agents on every surveilled host communicate periodically with a collection depot that is arbitrarily located. Agents are small, tight, invisible, tamper-resistant and low-load. Any touch whatsoever of local data is captured at the innermost operating system levels. Agents do 20,000-to-1 continuous log reduction, compress and encrypt bundles of these results, and push them to the collection system with end-to-end assurance, adapting to intermittent connectivity without intervention.

Because of Digital Guardian’s ability to capture information in real time, questions requiring full enumeration of past actions (i.e., prove no one outside the CFO’s staff read a certain document) and goals requiring zero-prep reaction (i.e., application whitelists and zero-day defense) become trivially feasible. At the same time, forensics becomes possible at near-zero reconstruction cost. Communities of trust become enforceable irrespective of conventional perimeters. The monitoring of information use at any level of granularity becomes auditably trivial and trivially auditable. Silent alarms can signal enforcement authorities of anticipated events or of unanticipated exceptions, and honest people can be coached to remain honest without the risk of inadvertently preventing anyone from getting their job done.

Digital Guardian is relevant in this day and age because an enemy able to strike from any location and without self-revelation commands the defender to focus on preemptive strategies. Preemption requires intelligence, and intelligence requires surveillance. For the electronic sphere, that surveillance has as its primary unit of observation either a data object or a person; only the former is at once versatile, no-load, inescapable, and an enabler of economic benefits that justify its existence.

It is, of course, possible to make a reference monitor fail by adding bells and whistles until it collapses under its own weight, just like any other computer tool. Digital Guardian does not have this disease, but alternatives do. Forcing intrusion detection to expand from detecting the illegitimate acquisition of legitimate authority to detecting the illegitimate use of legitimate authority is one example. Making stateful firewalls anticipate the effect of remote procedure calls (SOAP) passing through them by modeling the interior execution environment is another. Making patch management real time is another still.

Digital Guardian is relevant in this day and age because an enemy able to strike from any location and without self-revelation commands the defender to focus on preemptive strategies.

The total cost of ownership of any solution is dominated by its management, not its purchase. The management of a reference monitor is much more cost-effective than any single alternative – much more so than the accumulated sum of various highly-specific, non-platform products.

Being practical means keeping it simple. The central self-imposed limitation that jointly optimizes economic cost-effectiveness and security function is to focus security on data protection at the point of use. In truth, while this is an enormous simplification, there is little limitation involved. The total cost of ownership of any solution is dominated by its management, not its purchase. The management of a reference monitor is much more cost-effective than any single alternative – much more so than the accumulated sum of various highly-specific, non-platform products.

On the Adoption Curve

Because data has value and data is mobile, the handling of data drives regulation from this point forward. Because regulation, when operationalized, mobilizes the very things that security technology provides, it in turn drives security from this point forward.

Because an increasing threat means a shrinking defensive perimeter, wise companies make the point of use the focus of their protection efforts. Because the perimeter has contracted to the point of use, the host-based approach to data security is the only one that can be effective. Because regulation tends to demand categorical statements about what did and did not happen, only a platform approach deployed across the enterprise will do. Because the costs of host-based platform approaches are only now proportionally favorable to ignoring the threats, simplicity will, as always, be the hallmark of the security regime that works. Because some enterprises understand risk more than others, the adoption curve has already begun.

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

¹ The art and science of hiding information by embedding messages within other, seemingly harmless messages. <http://www.webopedia.com/TERM/S/steganography.html>

² “Reports that say that something hasn’t happened are always interesting to me, because as we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns; the ones we don’t know we don’t know.” Donald Rumsfeld, DoD, February 2002

³ sickness of the season

⁴ See <http://www.garlic.com/~lynn/secgloss.htm#12438> referring to <http://www.ietf.org/rfc/rfc2828.txt?number=2828>

ABOUT VERDASYS

Verdasys delivers solutions for Data Security at the Point of Use that address business requirements for assuring compliance with governance and privacy regulations; safeguarding the privacy of client and patient data; and preventing the misuse and theft of intellectual property. Offering compelling solutions for information audit and data containment challenges, the Verdasys Digital Guardian platform enables ongoing, comprehensive data-level monitoring and autonomous, real-time policy enforcement over enterprise information use.

VERDASYS.

950 WINTER STREET, SUITE 2600 WALTHAM, MA 02451

781-788-8180 TEL 781-788-8188 FAX

INFO@VERDASYS.COM WWW.VERDASYS.COM