Next Generation Web Application Firewalls: NG-WAF

White Paper

This paper describes Imperva's vision for the next generation of WAFs. It details Web application security problems and solutions today, and gives perspectives on the future. While this paper is not product specific, it indentifies areas where Imperva SecureSphere currently provides NG-WAF capabilities.

Table of Contents

Executive Summary

Security is often compared to a football game where to succeed the defense must be able to quickly adapt, outrun, and outplay the offense. The threat landscape has evolved, and attackers, the offense, have become more industrialized with greater organization, funding, focus and automation capabilities. This has been dubbed "The Industrialization of Hacking."

In addition to the industrialization of hacking are increases in threats from within – insiders – manifesting in sabotage, fraud, information leakage and risky business practices. Irrespective of attacks sourced from outside or inside an organization, point-and-click hackers or nation-states, attackers are targeting sensitive data, and that data is most commonly accessed through Web applications.

These applications are used for a variety of tasks ranging from customer self service portals, business-to-business communications, and online commerce to healthcare, critical infrastructure, and social networking. Unfortunately, most security experts agree that Web application security has been grossly lacking and the defense has been outmatched.

According to Jeremiah Grossman, founder and CTO of Web application security company, WhiteHat Security, "*The narrow pursuit of new Web application capabilities has created a complex landscape that most organizations lack the ability to secure without purpose-built Web application security solutions. Without a doubt organizations are fighting today's data-centric war with yesterday's network-centric approach resulting in application security preparedness a decade behind network security.*"

- Starting in 2003, in a series of incidents dubbed Titan Rain there were several attacks on US government agencies and defense contractors. These were thought to be Chinese in origin, although their precise nature (i.e., state-sponsored espionage, corporate espionage, or random hacker attacks) remain unknown. These attacks were responsible for the theft of sensitive information equivalent in size to the amount of data housed by the Library of Congress. Later in 2009, details on the Pentagon's $300 billion Joint Strike Fighter Project for the F-35 Fighter were stolen.
- In 2009, Heartland – which process over 100 million credit card transactions per month, experienced one of the largest data breaches in history. In 2010 Heartland settled with Visa by creating a $60 million fund for reimbursing affected card issuers.

Over the last decade cyber attacks have cost organizations millions and in some cases billions of dollars. These organizations have suffered brand damage, lost customers, reduced revenue, fines, and lawsuits. They have had to invest in victim notification and credit card monitoring services for their customers, public relations, and in some cases have gone out of business due to competitive pressure directly brought upon by the theft of intellectual property. From a government and military perspective, these attacks yield a far greater concern than industrial espionage, possibly impacting national security.

In response, Web Application Firewalls or WAFs rapidly evolved to defend Web applications against increasingly sophisticated attacks perpetrated by a growing number of attackers. WAFs became a bridge between organizational groups with different skill sets such as application developers and security operations so that risks could be better understood and mitigated. WAFs became easier to deploy and maintain while supporting various architecture types.

This paper will explore Imperva's vision of next generation WAFs, or NG-WAF in three interrelated sections covering: industrialized attack mitigation, interoperability and service delivery models, and risk management. It will also highlight some of the capabilities currently being delivered through Imperva's SecureSphere solution. But before we get there, we must understand what problems and solutions got us here.

Network-centric Attacks and Worms

In the late 1990s and early 2000s Hollywood loved "hacker" movies with the release of Sneakers, The Net, and Hackers. It was also the era of high-profile worms.

- The Code Red vulnerability was discovered in June 18th 2001. Within 48 hours Microsoft had a patch. Exploits didn't start until July 12th, 2001 against the un-patched systems. The estimated damage that Code Red left is over $1.2 Billion.
- Nimda – which is admin spelled backwards – launched later in 2001 and within 24 hours infected 2.2M systems causing over $500 Million in damage.
- SQL Slammer launched in 2003. It doubled its infection rate every 8.5 seconds; within 10 minutes 90% of all vulnerable, Internet accessible systems were compromised. SQL Slammer was considered the first Warhol worm, name after Andy Warhol's quote, "In the future everyone will have 15 minutes of fame." The idea was that in 15 minutes, attacks could impact all reachable targets on the Internet.

IPS

To address these worms and other network-centric attacks Intrusion Prevention Systems or IPS solutions emerged.

IPS were standalone solutions with limited capabilities. They were dependent on signature matching and creating blacklists of disallowed activity. While they proved to be effective for network-centric attacks and public attacks such as worms, they were most effective at preventing attacks where there were known signatures and exploits with known protocol vulnerabilities. These IPS solutions provided only a rudimentary line of defense against some of the first application-centric attacks such as SQL Injection and XSS (Cross-site Scripting). There was a critical need for a more robust solution that was effective against more targeted attacks at the application layer.

Targeted, Application-centric Attacks

The early 2000s marked a fundamental shift in attack types as attacks become more targeted. E-commerce became more common, more organizations started conducting business over the Internet, Web solutions were being deployed on a larger scale, and more applications were being created. This made it harder to find and fix vulnerabilities as well as schedule downtime to make these changes within these increasingly mission-critical applications. Even with the increased risk, organizations still wanted to stay focused on their core business, not developing secure Web applications. They demanded a cost effective solution with improved return on investment.

Regulations

In the years leading up to WAF, and directly following, there was notable government and industry reaction to cyber security. This came in the form of various regulations and mandates that WAFs were being depended on to help support through various protection mechanisms and for the demonstration of compliance to auditors.

Year	Industry & Government Reactions	Industry or Criteria
1995	European Privacy Law	Protects the privacy of individuals when their data is processed or transmitted
1996	HIPAA – Health Insurance Portability and Accountability Act	Healthcare
1996	Economic Espionage Act	Makes the theft or misappropriation of trade secrets involving commercial information, not classified or national defense information, a federal crime
1999	GLBA - Gramm-Leach-Bliley Act	Financial Services
2002	FISMA - Federal Information Security Management Act	US Federal Government
2002	SOX - Sarbanes-Oxley	Public Companies
2003	CA SB 1386 - California Senate Bill 1386 (40+ states have followed suit)	Requiring organizations that maintain personal information about individuals to inform those individuals if the security of their information is compromised
2004	Basel II	Financial Services
2006	PCI DSS – Payment Card Industry Data Security Standard	Companies processing credit card data
2006	NERC CIPS – North American Electric Reliability Corporation Critical Infrastructure Protection Standards	Electric Power
2008	Red Flags Rule	Financial Services
2009	HITECH - Health Information Technology for Economic and Clinical Health Act	Healthcare

WAF

As regulatory penalties became more real, and threats became more costly, organizations required a WAF solution more robust than IPS. The earliest WAFs, like early network firewalls were slow, obtrusive, and difficult to configure. Organizations demanded solutions that were easy to manage and deploy while being accurate and generating minimal latency. They also demanded a solution that could be deployed and used by security professionals without involving application developers. Interestingly, while there was a desire to deploy without application developers, WAFs were also seen as a bridge to bring these groups closer together for application risk evaluation and mitigation. For example, the security team could show the application developers exactly how the applications were actually being used, what attackers were targeting, and the impact they had, without being application developers themselves. This helped reduce the window of exposure – the amount of time between the discovery of an issue and its resolution.

Following are some of the hallmarks of WAF.

» **Whitelisting**: Applications are extremely dynamic, so WAFs had to be able to learn the application and then continually re-learn as code changes were made. WAFs couldn't be dependent upon binary logic commonly associated with IPS style blacklist signature matching which simply didn't scale to address various attack and evasion techniques such as Blindfolded SQL Injection, SQL Injection Signature Evasion, Parameter Tampering, Forceful Browsing, and Cookie Poisoning.

Imperva SecureSphere uses "Dynamic Profiling" to dynamically learn the structure, elements, and usage of Web applications giving it greater insight into how applications are actually being used. This process automates whitelist creation – by defining explicitly what is allowed. Because these lists can contain thousands of URLs, form fields, parameters and cookies, automation is critical.

» **User Session Reconciliation**: As Web sites became Web applications and were used to perform transactions, the user context in the Internet evolved. This required WAFs to adapt in order to improve accountability and set user context controls.

Imperva SecureSphere offers "User Tracking" to not only understand Web application transactions in terms of user accountability, but also apply accountability to SQL database transactions made by the application on behalf of the user. This directly addresses issues such as session pooling and a lack of accountability between applications and databases.

» **Architectural Choices**: Deployment flexibility beyond the software-based reverse proxies of early WAFs was necessary for organizations requiring transparent and high performance architectures.

Imperva SecureSphere uses a kernel-based inspection model atop a transparent layer-2 bridge which provides full application and session awareness without the cumbersome requirements of a proxy architecture. Additionally, Imperva SecureSphere supports other network configurations such as: router, transparent proxy, and non-inline monitor – i.e. span port or tap deployments. Because WAFs are being deployed as mission-critical solutions, Imperva SecureSphere also supports high availability architectures.

» **Centralized Management**: A single management system became a core theme for WAFs as organizations began deploying a larger number of WAFs which were in some cases globally distributed. This also allowed organizations to build, maintain, and enforce a unified security policy across their entire organization.

Imperva SecureSphere allows centralized management of multiple WAFs and related Imperva SecureSphere solutions such as database firewalling and Database Activity Monitoring (DAM).

» **WAF + Database Security:** Integration with database firewalls and DAM solutions started to emerge as organizations increasingly viewed Web applications and databases as part of a larger data security strategy.

Imperva SecureSphere allows organizations to track how users are truly interacting with their data through the Web application, into the database, and back. Imperva provides blocking and monitoring at application and database layers and is particularly useful for mitigating threats by privileged users such as malicious DBAs.

- *Many regulations are sensitive to how users are interacting with sensitive data. For example, Sarbanes-Oxley requires access to and modification of financial data to be tracked. Imperva gives organizations full visibility into how data flows in the organization. Within Imperva SecureSphere each Web transaction is mapped to the relevant queries and data storage objects. Based on this mapping, users can define and enforce various rules in the context of the data rather than the context of a certain transaction.*
- *Imperva SecureSphere also helps address data leakage by allowing organizations to control the type and amount of data that flows out of the organization through the Web application.*

» **Correlation**: Correlation helped pull everything together to detect stealthy attacks by not just processing each piece of data in a vacuum, but combining various sources to create a prioritized response.

Imperva SecureSphere leverages correlation across multiple applications and database thus minimizing false positives and false negatives by taking a much broader and unified view of user and system activity.

While WAFs are far superior to IPS for application-centric attacks, attackers aren't standing still. More sophisticated attacks are emerging with the promise of greater devastation. In particular, financial and political motives are ushering in a new era of funded and motivated attackers representing state-sponsored espionage, corporate espionage, organized crime, and for-profit cyber criminals. Industrialized hacking is becoming a reality.

Organizations require more expansive solutions for addressing these new industrialized attacks. This demand is driving WAF industry to take a major evolutionary leap forward.

Industrialized Attacks

Industrialization of Hacking

In a September 2009 piece titled "The Top Cyber Security Risks", the SANS Institute stated that attacks against Web applications constitute more than 60% of the total Internet attacks.

Hacking is a profitable business, and like any business, to achieve scalability, it must become organized and various groups must become specialized. At its core, industrialized hacking can be understood by looking at three areas that include the roles and responsibilities of the individuals involved, the optimization of compromised resources, and the automation of the attack vectors.

Roles and Responsibilities

- Vulnerability researchers and exploit developers keep their distance from the actual exploitation of systems
- Botnet growers focus on enlisting as many unsuspecting systems into their botnet army as possible and disabling other bots that may have previously compromised the target
- Attackers use the botnets and exploits of others
- Consumers receive the ill-gotten gains from these crimes

Optimization of Compromised Resources

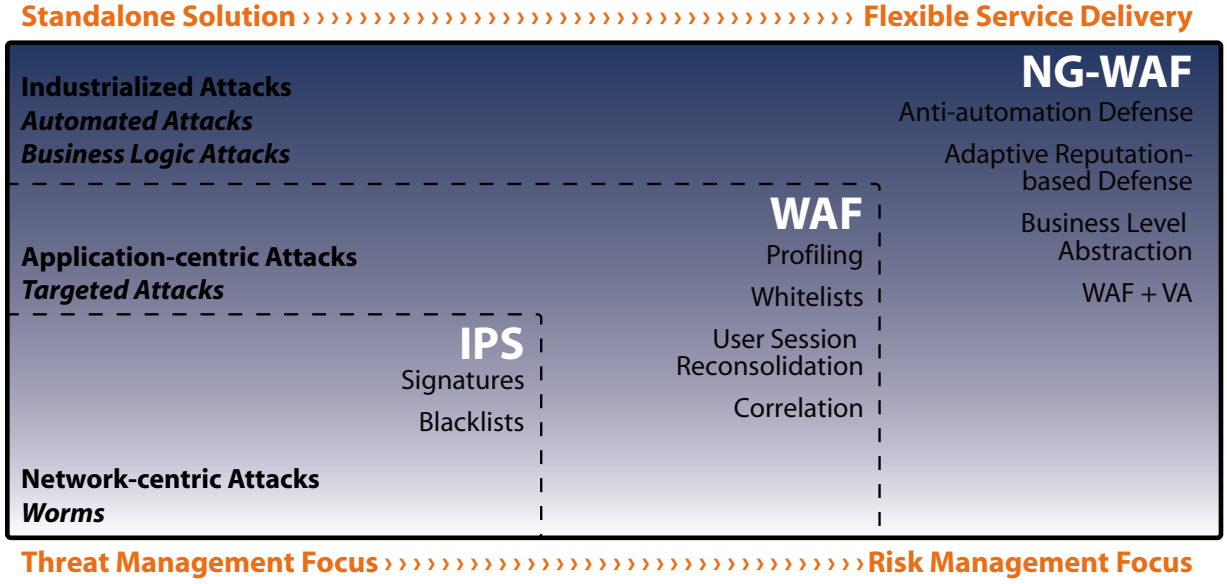
- Directly stealing intellectual property, personally identifiable information, financial data, etc from the compromised system
- Leveraging the compromised system as a member of a botnet army for malware distribution, DDoS, spam, Phishing, and other nefarious acts

Automation of Attack Vector

- System exploitation and botnet growth must be automated to scale
- Targets selection will leverage search engines to increase efficiencies and effectiveness
- Kits and templates exist so that operational costs can be minimized and ROI improved
- Botnets can be centrally managed or leased as a service

NG-WAF

This illustration depicts the evolution from IPS to WAF to NG-WAF. Each incarnation incorporates and improves on the last.



IPS was very focused on network-centric attacks and worms. It was efficient in this capacity, but signature-based matching i.e. blacklists, made it ineffective for addressing application-centric attacks. Additionally, it was a standalone solution with only a threat management focus. WAFs were more dynamic and added whitelisting, and perhaps most importantly were designed to specifically address application-centric attacks. However, like IPS, they were standalone solutions with only a threat management focus. Imperva's vision of NG-WAF is such that it will build atop IPS and WAF and further the application-centric attack mitigation capabilities to include industrialized attacks. Unlike IPS and WAF however, NG-WAF will be interoperable with other solutions such as Vulnerability Assessment (VA), offer flexible delivery such as Managed Security Service Provider (MSSP) and cloud-based models, and will go beyond threat management to also include risk management.

Perhaps no attack type is a stronger requirement for WAFs to evolve than automated attacks.

Industrialized Attack Mitigation

Automated Attack Mitigation: Fight Bots & Other People-less Attacks

It's about time law enforcement got as organized as organized crime.
- Rudy Giuliani, former New York City Mayor

Automated Attacks

More attacks are being discovered that aren't perpetrated by a single individual or from a single system, but rather by an organized network of zombies or bots operated by a single command and control center. These botnets may be developed and used by a single attacker, organized group, or perhaps leased out to other individuals or organizations to target a specific company, send spam, conduct DDoS, RFI probing, SQL injections, propagate malware, and any number of malicious acts.

2008 and 2009 brought an increasing number of such automated attacks. Every day, millions of bots are unwittingly enlisted into service. The attack vectors for the bots have also been maturing with the addition of business logic attacks such as brute force logins, comment spam, and click fraud. Detection is very difficult because automated attacks have separate transactions that alone appear legitimate. Only when viewed holistically is the bigger picture seen. Early industry attempts at stopping these attacks with simple IP blocking failed because these attacks were often relayed through any number of Web proxies and the origins were always moving.

Anti-automation Defenses

Imperva SecureSphere currently offers anti-automation defenses and is able to leverage a number of techniques including passive rate measurement, and request structure analysis combined with proactive behavior fingerprinting to identify non-browser and non-human behavior – i.e. automated attacks.

Countermeasures are quite different for automated attacks as opposed to an attack by a single human. For example, if an attack is being conducted live by a single person, after a few blocking requests are initiated, the attacker they may stop and move on. In contrast, an automated attack may exhaustively search for usernames, passwords, and other resources, and it won't just stop because of a few blocked attempts. Think of a vertical network port scan searching for 65,000 possible open ports. Only a few will be active, but it will scan for every possibility no matter how many ports are closed.

With automated attacks, simply blocking may not be the best approach. Instead, slowing down the transaction with capabilities discussed in the next section might be more desirable since a legitimate request may be very difficult to distinguish from a malicious request. Ultimately, in any mitigation scenario, the desired impact is to have minimal impact on the legitimate user, while causing the attack to fail.

This last statement is easier said than done. It requires a previously untapped resource for Web application security solutions and will require NG-WAF capabilities to look beyond the organizations they are protecting.

Adaptive Reputation-based Defense: Know Thy Enemy

Know thy self, know thy enemy. A thousand battles, a thousand victories.
- Sun Tzu, The Art of War

Understanding the Attacker's Reputation

There was a time when application attack traffic was rather scarce, and the number of attack vectors was low. So it wouldn't be surprising to have just a few attacks on a Web application a day. However, with automated attacks, it doesn't matter if a site is high profile or obscure, contains sensitive data or doesn't. It's going to be attacked. As such, the number of alerts will be much higher for everyone, and this can lead to paralysis of analysis. Unfortunately, the more sophisticated and directed attacks that an organization might be interested applying further analysis to may be obscured by sheer volume.

Many attacks are scripted. This reduces the skill required to launch an attack and increases the number of attackers. More attackers, running more attacks equals more alerts. In fact, knowing this, some exploit developers will release their code publically so that attack scripts will be used by many, and can't be traced back to a single individual.

Attackers are leveraging global resources to discover vulnerabilities through processes like Google Hacking, and executing attacks on those systems with heightened anonymity on a large scale by leveraging botnets and anonymous proxies. In order to combat these attacks, organizations need to be able to leverage solutions that are plugged into these same global resources. By having timely, real-world information about attacker sources and attacker vectors, it is possible to more quickly determine how to address malicious traffic, and what warrants more detailed investigation. For example, is the attack sourced from: anonymous proxies, Tor IP addresses, known malicious address space or domains, geographies of interest, known botnets, or Phishing URLs.

Adaptive Defenses

Imperva SecureSphere offers adaptive defenses today called ThreatRadar. ThreatRadar aggregates information based on subscription service feeds and other sources that constantly monitor Internet activity on a global scale. This information is continually updated with attacker sources and attacker vectors.

This intelligence surrounding the attackers improves and automates attack detection from known malicious sources. It reduces risk by identifying malicious users before they execute attacks, such as in reconnaissance phases or after signs of suspicious activity. It helps detect denial of service attacks that may otherwise be difficult to differentiate from legitimate activity. Investigation times are reduced because security alerts will be able to clearly identify the known malicious sources.

Based on the ThreatRadar service, every organization benefits from reputational data gleaned from attacks around the world. This information is then transformed into security policies applied directly within Imperva SecureSphere. Based on the attacker and attack vector, multiple responses can be used including: attack blocking, alerting, re-direction, presentation of multi-factor authentication choices, and challenge-response such as CAPTCHA.

Understanding attackers better is also covered in the Building Security In Maturity Model (BSIMM) created by Dr. Gary McGraw, Brian Chess, and Sammy Migues. BSIMM is a real-world set of software security activities organized so organizations can determine their software security initiatives compared to others and how it should be updated. Sections from the BSIMM's Intelligence Attack Models highlight the need for:

- Identification of potential attackers in order to understand their motivations and capabilities
- Collection and publication of attack scenarios
- Gathering of attacker intelligence
- Building attack patterns and abuse cases

Automated attack mitigation and adaptive reputation-based defense can also be combined when mitigation business logic attacks.

Business Logic Attack Mitigation: Understand the Business to Protect It

Obviously crime pays, or there'd be no crime.
- G. Gordon Liddy, masterminded the first break-in of the Democratic National Committee headquarters – i.e. Watergate

Attacking Business Logic

Just because defensive mechanisms have matured for protecting against traditional technical attacks, doesn't mean that attackers are sitting still. Attackers are still well motivated to find new methods of compromise. Business logic attacks target the logic of a business application, not a technical vulnerability. As opposed to "traditional", technical, application attacks, such as XSS or SQL Injection, business logic attacks do not contain malformed requests and include legitimate input values making them difficult to detect. Business logic attacks abuse the functionality of the application, attacking the business directly. These attacks can be further enhanced when combined with automation where botnets are used to challenge the business application.

In an OWASP (Open Web Application Security Project) article titled Testing for Business Logic, the natures of these attacks are outlined.

> *Business logic can have security flaws that allow a user to do something that isn't allowed by the business. Frequently, these business logic checks simply are not present in the application.*

Jeremiah Grossman, founder and CTO of WhiteHat Security published a whitepaper on this topic titled, Seven Business Logic Flaws That Put Your Website at Risk. In an excerpt from that paper he specifically calls out CAPTCHA, a method discussed earlier, as a part of the solution. Jeremiah states, "As an alternative to an account lockout, a CAPTCHA system may be employed if an account has received too many failed login attempts. This method has the benefit of preventing brute force attacks, without the potential side effect of locking out legitimate users."

Consider a business logic attack on a Web application designed to allow customers to buy concert tickets on-line. The purchase flow is designed to prevent customers from paying for the same seat, so once a user selects their seat, it is reserved until they close the application or buy the ticket.

A customer could exploit this application by purchasing one ticket and then initiating the purchase process for all remaining seats without entering billing information or closing the application. This attack results in all seats being reserved and blocking others from purchasing the remaining seats. The attacker is then free to buy more tickets and scalp them for a profit.

This is an example of the software doing what it was intended to do. A code review, vulnerability assessment scanner, or traffic inspection solution alone would likely be ineffective because such processes and solutions rarely consider what business operations are associated with transactions making application activity and network traffic appear legitimate. Internal applications also implement a great deal of business logic which could potentially expose them to attacks from malicious insiders. As such, security controls need to consider internally and externally-facing applications.

Leveraging Business Level Abstraction

It is Imperva's vision that for NG-WAFs to address business logic attacks they first need to have an understanding of business operations – i.e. a business level abstraction, atop the technical-centric capabilities they provide. Web applications communicate using HTTP, but this is simply a method to implement complex business transactions. The technology and the business intelligence have been decoupled in application security solutions thus far, making it extremely difficult to detect and prevent business logic attacks.

NG-WAFs will have to enrich various technical application elements by mapping them to business transactions and applying security policies based on that mapping. That will help in the detection of business logic attacks. Since these attacks look like legitimate traffic in terms of structure with perhaps the exception of things like rate, flow, or related characteristics that are not part of the Web request itself, traditional WAF technical detection techniques won't be enough, thus a business abstraction is imperative. Business level abstraction of Web traffic and security rules will couple the necessary variables to effectively address these complex tasks.

Up to this point, the focus has been on NG-WAF as a standalone solution. Managing risk however has created a need for NG-WAFs to expand beyond standalone solutions to integrate with vulnerability assessment solutions for improved patch management. Similarly, as organizations continue to seek new and innovative models for Web application delivery, NG-WAF solutions will need to operate within MSSP and Cloud environments.

Interoperability and Flexible Service Delivery Models

Vulnerability Assessment and Patching: The Bad Guys Know Your Vulnerabilities, Shouldn't You?

Security is always going to be a cat and mouse game because there'll be people out there that are hunting for the zero day award, you have people that don't have configuration management, don't have vulnerability management, don't have patch management.
- Kevin Mitnick, convicted computer hacker

Patch Management Issues

Deven Bhatt, the chief security officer of ARC, stated, "PCI recommended code review or WAF, but we found the 24x7 protection offered by a WAF a particularly valuable asset. For instance, a WAF can be used to protect all Web applications, while most people would only perform code review on applications that are directly under PCI regulation, such as payment processing applications. This would leave the other Web applications wide open to attacks."

WAF is bad; just write better code. Code is perpetually broken; just use a WAF. It's hard to believe today, but this was once a heated industry debate. While the argument is all but dead, this bifurcation between the WAF community and the application development community once made organizations feel that they had to choose WAF or code review and there was no middle ground. Much of this false dichotomy was brought upon by PCI DSS requirement 6.6 which originally sited that WAF "or" code review was needed to achieve compliance: enter the battle for budget. The requirement has since been amended to support defense in depth with a combination of good programming practices, "and" WAF. PCI DSS 6.6 Supplement now states, "Proper implementation of both options (application code review and Web Application Firewalls) will provide the best multi-layered defense." It would be hard to argue with this logic as the notion of defense-in-depth on the network side – scan for vulnerabilities, deploy devices with hardened security, and use controls like network firewalls is a foregone conclusion.

Prudent use of Microsoft's SDL, BSIMM, OWASP CLASP, and others can dramatically improve the quality of software, and the security of the information it's processing. This is especially true to the point where flaws are not interfering with common usage of the software and vulnerabilities are not abounding. So yes, write better code, but don't be under any suspicions that it is or will ever be perfect.

Ultimately, Web application vulnerabilities should be patched, but there are a number of reasons why this doesn't always happen in the real world.

- The Web application patch – either proprietary or commercial -- needs to be created.
- Changes need to go through QA.
- A freeze period, other changes, or inability to make a business case for the change creates delay.

In January of 2002 a now infamous memo was leaked out of Microsoft. The author was then CEO Bill Gates, and the subject of the memo was simply Trustworthy Computing. Following is a snippet from that memo where he outlined the need for enhanced security.

> *"Today, in the developed world, we do not worry about electricity and water services being available. With telephony, we rely both on its availability and its security for conducting highly confidential business transactions without worrying that information about who we call or what we say will be compromised. Computing falls well short of this, ranging from the individual user who isn't willing to add a new application because it might destabilize their system, to a corporation that moves slowly to embrace e-business because today's platforms don't make the grade."*

Over the last few years Microsoft has made tremendous improvements in the security of their software applications and operating systems. Much of this can be attributed to Microsoft's mature Security Development Lifecycle (SDL). However, on October 13th 2009 they released the largest security patch in the company's history. Not because they are getting it wrong, but because they have millions of lines of code, and there are inherent limitations to the SDL as the first and last line of defense.

Some would claim that Microsoft had just reached the inherent limits of real world software debugging processes. The law of big numbers, applied to lines of code, gives us a non-zero prediction as to the number of software flaws per 1000 lines of code. A mathematical postulate shows that guaranteeing the correctness of a general computer program is a non-decisive problem – i.e. it cannot be solved in a finite time. In fact there is a point in time in which any increase in QA resources and time has a negligible effect over software quality.

Virtual Patching

In an August 2009 blog post, Neil MacDonald from Gartner discusses the convergence of VA and WAF in a piece titled - Security No-Brainer #9: Application Vulnerability Scanners Should Communicate with Application Firewalls. He states, "It's time to start requiring this capability [WAF + VA] in our web application security testing tool providers via partnerships with web application firewall vendors.

While many organizations are familiar with network and operating system vulnerability assessments, and patch management, there exists another layer – Web application VA. This type of VA goes beyond finding open ports, services with clear text protocols, and out-of-date encryption packages to actually finding vulnerabilities within the application itself. Leveraging VA with NG-WAF will improve both WAF and VA capabilities. This combination will improve the security life cycle, allowing for more effective and coordinated responses to security issues, and by doing so reducing overall risk.

According to Marc Appelbaum, Manager of Information Security at Vonage, "WAF is always on. It's constant application security, whereas other things, such as code reviews, or even vulnerability scans, are point-in-time snapshots of an environment. The WAF is always looking at the traffic that's traversing it. It's always monitoring for vulnerabilities.

Today Imperva SecureSphere interoperates with VA solutions that periodically reassess new types of attacks, application changes, and configuration changes, and export that detail into Imperva SecureSphere for "virtual patching." Virtual patching is the ability to apply Web application patches on the NG-WAF without actually making changes directly on the Web application.

One disadvantage of a VA tool alone is coverage and the automated crawling process encountering problems within covering complex applications. It's Imperva's vision that the NG-WAF will be able to discover Web applications and export details into the VA solution, in addition to preventing attacks aimed at exploiting those vulnerabilities, monitoring user interaction, and continually learning how the applications operate. By exporting profile information generated by the NG-WAF inspecting real user requests and the introduction of new application modules the VA solution could launch a new scan allowing greater coverage and a more exacting assessment. The relationship between NG-WAF and VA also means that once vulnerabilities are discovered, the VA's output will be efficiently and effectively used by the NG-WAF as a security policy that will be able to alert on or block attempted exploits.

Imperva further hypothesizes that NG-WAFs will notify the VA solution that the Web application has been virtually patched; the VA solution will re-scan the Web application ensuring the vulnerability is no longer available. Additionally, the NG-WAF will be able to notify the VA solution about specific application changes thus enabling a rapid and targeted scan of the new changes without the overhead of crawling the entire application and scanning the application in its entirety. This will have an added benefit of reducing risk in terms of time between scans associated with application changes on larger applications that may only periodically assessed.

NG-WAF and VA will still focus on their core competencies, but their synergies will reduce risk and limit the window of exposure. Operationally, managing the vulnerabilities also becomes more formalized as NG-WAF solutions start providing a centralized framework for tracking, managing, and mitigating Web application vulnerabilities.

For example, if a new vulnerability in a certain PHP module has been disclosed, the security team can mitigate that vulnerability through integration with vulnerability assessment results associated with the NG-WAF policy or possibly with an automated application security update. Imperva sees this solution going further and integrating with static code analysis and providing links between exploits and bugs thus closing the loop between development and production environments. This will truly give developers a view into how attackers are trying to exploit vulnerabilities, and how the applications are "really" being used as opposed to how they were designed to be used.

No two environments are cookie cutter. Variances abound in IT, especially around Web applications. Because of these variances, the need to manage risk and the need to mitigate threats, many organizations are leveraging alternative delivery models offered through MSSPs and Cloud services.

MSSP and Cloud Computing Delivery Models: Your Solution Your Delivery Platform Choice

Cloud computing will be as influential as e-business.
- Gartner

Focusing on Core Business Initiatives

Organizations today want choice. And they want to make choices that support their core business initiatives. Businesses of all sizes desire to reduce costs and generate value for their customers. The growing risk of Web application attacks along with new regulatory requirements has underscored the need for Web application security. Some businesses are seeking alternative delivery models such as MSSPs.

MSSPs

Imperva already works with a number of MSSPs that are protecting customer Web applications with Imperva SecureSphere. Further, Imperva has deployed its own MSSP offering to support customers and partners that may not have application security resources themselves.

Imperva SecureSphere is extensible enough to operate within MSSP environments and delivers capabilities specifically designed for the MSSP.

MSSP core requirements:

- Most WAF solutions can have several WAF instances reporting to a single WAF manager; MSSPs will require a NG-WAF manager of managers in order to build a scalable hierarchy
- High-availability and disaster recovery capabilities will be essential to the core NG-WAF architecture
- NG-WAFs will need customizable reporting capabilities that will help MSSPs deliver personalized customer service without sacrificing operational scalability

Taking Advantage of Greater Scalability with Reduced Infrastructure

Some organizations are completely virtualizing their approach to online business applications. This can apply to very large organizations that want the advantages and scale that cloud computing models can deliver, but it also fits within smaller organizations. Often these smaller organizations turn to virtualized delivery models as a way to avoid infrastructure investments and leverage the IT expertise of the service providers rather than building in-house capabilities. Regardless of the size of the customer using the cloud service, they will still require Web application security.

Cloud Computing

NG-WAF vendors again will have to morph their delivery models to consider cloud computing.

Imperva has a number of cloud computing partners providing Web application security today to their customers via Imperva SecureSphere. These partners leverage Imperva in a SaaS model without impacting their network architecture or needing additional security staff to operate. This delivery model focuses on minimizing the time required for implementation, reduction of operational costs, and providing security with minimal effort. Additionally, Imperva SecureSphere can protect on-demand instances of applications and accommodate fluctuation in Web application load.

Regardless of solutions being applied directly or through services, to effectively mitigate industrialized attacks and manage risk, organizations must look inwards to more accurately understand the targets.

Risk Management

Application Discovery: You Can't Protect It Unless You Know About It

There are known knowns. These are things we know that we know. There are known unknowns. That is to say, there are things that we know we don't know. But there are also unknown unknowns. There are things we don't know we don't know.

- Donald Rumsfeld, Secretary of Defense under President George W. Bush

Knowing Where the Targets Are

Simply knowing where Web applications are and their details has been a difficult issue for many years because the right automated tools to augment the operations staff didn't exist. This process is actually getting harder to mange because of several factors including organization expansion, mergers and acquisitions, and more frequent department restructuring, not to mention the prolific growth in Web applications in general. In some cases it's not even the production Web applications that put organizations at risk, but rather test systems that may contain sensitive data, grant unauthorized access, or have embedded details valuable to attackers such as passwords, comment fields, or detailed error messages exposing system variables. NG-WAFs will be able to proactively and reactively address this need and provide a more accurate mapping of organizational applications.

Proactive Application Discovery

Imperva's vision is that in proactive modes, the NG-WAFs will be able to periodically scan networks for Web applications. In addition to discovering the Web applications, they will be able to understand their vulnerabilities as discussed earlier. Another NG-WAF capability will be Web application change tracking. NG-WAFs will be able to determine if there has been some form of unapproved Web application change such as a defacement attack. By evaluating Web application files against known good files, the NG-WAF will be able to detect and alert on changes.

For databases, Imperva SecureSphere already offers this capability of proactively discovering databases, the sensitive data they contain, and classifying that data.

Reactive Application Discovery

Imperva's vision also extends to reactive models, where the NG-WAFs will be able to continually monitor live Web traffic. Based on traffic analysis, the NG-WAF will be able to discover new application modules, detect security issues such as clear text passwords, and the use of improper output encoding. By detecting these issues in a reactive mode, they can be addressed before they impact an organization's Web applications, customers, and reputation.

This combination of discovering Web applications through proactive and reactive modes is necessary for knowing what you are protecting, application changes on those assets, and how to adjust your defenses. Proactive or reactive alone will not address this need. By further combining timely Web application vulnerability assessment data with all known Web applications, the NG-WAF will have a baseline of all the potential attacker targets and the vulnerabilities on those targets. This will result in more exacting preventative controls and responses.

Within the critical infrastructure industries there is a notion of "survivability." This idea is that successful attacks are going to happen – regardless. So they want to ensure two things. First, that irrespective of the attack, services such as water, electricity, and transportation remain on. Second, following that attack, there is enough forensic evidence to analyze and build in new security controls so it doesn't happen again. This is why audit information is such a critical part of any security strategy.

Web Auditing: Following the "Webprints"

When you have eliminated the impossible, whatever remains, however improbable, must be the truth?
- Sherlock Homes, The Sign of Four

Issues with Forensic Data

Forensic insight derived from audit information into what happened during an incident is expected to become an increasingly important theme within regulations. Beyond regulations, audit can be as valuable as a security solution. Consider a malicious insider. During an investigation there are several questions – was this person in fact malicious, or were their actions simply careless. Auditing helps address this question along with other pressing points such as – how long have they been doing this, what else have they been doing, and who else might be doing something similar. Unfortunately Web application auditing is often not enabled because of system resource constraints. If the attacker is a privileged user, they may be able to modify the audit information making it unreliable. Often, when auditing is turned on, it is done so in a minimal way that generates incomplete audit trails that are useless for actual investigations. These points taken collectively have lead to organizations operating with little to no usable Web application audit information.

Reliable and Complete Forensic Data

Imperva's vision is that NG-WAF solutions will need to audit usage trends, bandwidth utilization, and performance levels in addition to security data. They will need to provide visibility into Web transactions from the application level and at the business abstraction level discussed earlier. In doing so, they will be able to provide insight into user activity, Web business transactions and application usage. To optimize these capabilities, they'll need to make use of analytic tools for processing and analyzing the mountains of data that busy Web applications generate and give analysts an intuitive, visual interface to expedite analysis. The business abstraction components will express the relationships between HTTP traffic and the business transactions they support. This will render greater relevance and understanding for those not intimately familiar with application programming, but are aware of essential business processes.

Detailed investigations will be able to yield patterns, anomalies and behavioral interactions that can then in turn be used to generate more exacting security policies, reports, and remediation efforts. Other key capabilities associated with Web auditing will be analogous to the features found in Imperva's DAM solutions.

Imperva SecureSphere currently provides database auditing that resides outside of the database and captures bi-directional communication between users and the audited system. Because the database no longer needs to be auditing, there aren't concerns over negative performance impact, Separation of Duties (SOD) between security analysts and privileged users such as DBAs, or concerns about the right level of auditing being enabled. Finally, all the litigation-quality data will be stored centrally regardless of the distributed nature of the various databases or their heterogeneity.

Another capability that auditing should allow NG-WAFs to provide is virtual change management. This is the notion of testing changes to the Web application environment virtually by replaying captured audit information through the WAF to determine how such changes would impact the real systems. With this capability, more diligent testing could be done as part of the change management process to determine if the changes will create any adverse or unexpected issues, before changes go into production.

Conclusion

This paper has explored Imperva's vision for NG-WAF as well as the threats and organizational requirements driving that vision. Where applicable, offerings that are available today through Imperva SecureSphere have been highlighted.

Threats have evolved and become industrialized. Attackers are targeting Web applications to get data. The criticality of data has evolved. Data drives businesses more today than at any other time in history. In order to protect the business organizations need to protect the Web applications and the data. This protection requires the next generation of WAF.

Imperva
Headquarters
3400 Bridge Parkway
Suite 101
Redwood Shores, CA 94065
Tel: +1-650-345-9000
Fax: +1-650-345-9004

Toll Free (U.S. only): +1-866-926-4678
www.imperva.com

#WP-NG-WAF-0210rev1