Improving Web Application Security:

The IBM® Edge Delivery Services Approach to WAF

IBM

Table of Contents

Introduction

Many security professionals consider the web application firewall (WAF) to be among the most complex security technologies on the market today. Sitting in the middle of the HTTP conversation between users and a web application, the WAF inspects HTTP traffic passing through it for any attacks as defined by a list of rules. The complexity of this task comes inherently with its basic definition, that of:

- Relying on a pre-defined list of rules to identify malicious HTTP requests, with thousands of potential exploits to guard against. In addition, new attack vectors or additional permutations of existing ones are continuously being discovered and exploited.
- Relying on a pre-defined list of rules to identify malicious HTTP requests interspersed with legitimate HTTP traffic, while the characteristics of legitimate traffic differ on a per-application basis and change over time.

Complicating this task, organizations have little ability today to measure, understand or quantify the effectiveness of their WAF solution in a real-time and unpredictable environment. This has led to challenges experienced by organizations in terms of accuracy, performance and management overhead. The IBM Edge Delivery Services approach to WAF combines a) an anomaly detection model with b) a repeatable testing framework to measure effectiveness, c) threat intelligence to identify the latest threats, d) a cloud platform for global scale, and e) managed security services to help organizations better protect their websites and web applications over time.

Challenges with Deploying WAFs

Many organizations have web application firewall solutions deployed today. However, these deployments often fail to meet their initial expectations in terms of effectiveness, ease of management and impact on protected web applications. In a March 2015 report, the Ponemon Institute surveyed 594 IT professionals responsible for web application security about the status of their organization's WAF deployment.

While 68% of the respondents had deployed a WAF, only 20% had deployed it inline. Twenty-three percent had deployed their WAF out-of-line, while a further 25% deployed it in a combination of inline and out-of-line configurations. Because a WAF must be deployed inline in order to block malicious requests, this behavior indicates that organizations face significant challenges in deploying their existing WAF solutions, including:

Breakdown of WAF deployment
Ponemon Institute, January 2015

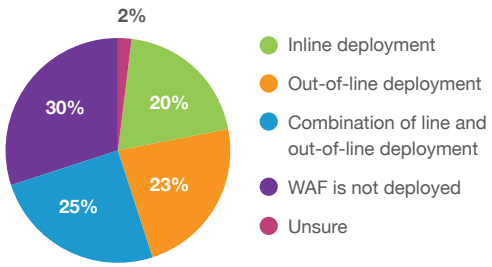


Figure 1: Response to the question, "What best describes your organization's approach to WAF?"

- **Accuracy** – most WAF vendors either do not provide accuracy measurements or measure it using a limited set of test traffic, and organizations often do not understand the potential impact of false positives or false negatives until after they have purchased a solution. As a result, organizations often purchase a WAF intending to deploy it inline, but deploy it out-of-line once the impact on legitimate users is discovered.
- **Not enough people** – many organizations underestimate the overhead required to properly maintain a WAF over time. Without dedicated resources, WAF configurations can quickly become out of date as applications change, risking a higher rate of false positives and false negatives. As a result, organizations often pull their WAF out of line once the configuration becomes out of date.
- **Performance** – a WAF solution with insufficient scale can reduce the performance of the protected applications. Because of the challenge in quantifying any performance impact prior to purchasing a solution and deploying it inline, organizations often purchase a WAF solution with the intention of deploying it inline, but end up deploying it out of line once they observe a web performance impact.
- **Fail-open/fail-closed** – under extremely high traffic conditions, traditional hardware WAF appliances are designed either to fail-open or fail-closed. In a fail-open situation, the WAF will allow all traffic to the application without inspecting for or blocking malicious traffic. In a fail-closed situation, the WAF will block all traffic from the web application. Because of this behavior, organizations often need to pull WAF appliances out of line during extremely high traffic in order to maintain web application availability.

WAF Design Principles

In the Ponemon survey, the low percentage of respondents with an inline WAF deployment points at a significant industry challenge. Any vendor can build a WAF solution and bring it to market with relative ease, as demonstrated by the prevalence of commercial offerings built around the open-source OWASP ModSecurity® Core Rule Set (CRS). However, it is very difficult for a vendor to design an effective WAF – one that can be deployed inline to protect organizations' applications over time as new vulnerabilities are discovered, the amount of protected web traffic grows and the web applications themselves change.

What factors contribute to the effectiveness of a WAF? While no security solution can be 100% effective, all WAF solutions should strive to provide the following:

1. **Accurate protection** – can it stop more web attacks while blocking fewer legitimate users?
2. **Visibility into attacks** – can it remove the guesswork from identifying and responding to attacks?
3. **Adaptability to changing threats** – how well will it stop unknown attacks?
4. **Adequate scale** – can it handle all of the web traffic that an application is likely to see, without becoming a bottleneck?
5. **Ease of management** – how much effort is required to deploy and manage it over time?

Accurate Protection

Every WAF solution relies on the quality of its rule set to identify web attacks without blocking legitimate users. Historically, WAF solutions have required organizations to make a tradeoff between false positives and false negatives – typically prioritizing the minimization of false positives at the expense of a greater number of false negatives. While this alleviates many organizations' concerns about accidentally blocking legitimate users, it also protects against fewer web attacks. A more effective approach offers a lower rate of both false positives and false negatives, increasing the accuracy of the protection provided while still minimizing impact on legitimate users.

Understanding Accuracy

Accuracy measures the ability of a WAF to simultaneously stop attacks while not inadvertently blocking legitimate users and considers four variables:

- **True positives (TP)** – real attacks that are properly identified and blocked by the WAF.
- **False positives (FP)** – legitimate user requests that are improperly identified as an attack and blocked by the WAF.
- **True negatives (TN)** – legitimate user requests that are passed through to the application.
- **False negatives (FN)** – real attacks that are not properly identified and blocked by the WAF and are passed through to the application.

Visibility into Attacks

Traditional WAF solutions provide a never-ending stream of alerts and rely on administrators to analyze the alerts and determine if an attack has occurred. This requires web security resources and expertise that many organizations do not have. A more effective approach provides visibility into and context around online attacks that have occurred – notifying an organization when, where and how an attack occurred and immediately providing administrators with any pertinent information. This relieves administrators of the burden of determining whether or not an attack has occurred and instead enables them to immediately focus on any additional response, if needed.

Adaptability to Changing Threats

Organizations must continuously update their WAF solution to address new vulnerabilities as they are discovered. In this context, most WAF solutions focus on how quickly a new rule can be created and deployed when needed. However, this ignores two other requirements that must first be met:

- **Awareness of vulnerability** – most organizations do not have visibility into the latest threats and must rely on their security vendor. However, WAF vendors often do not have the visibility themselves and struggle to notify customers of new attack vectors or provide rule updates in a timely manner.

- **Security resources and expertise** – most WAF vendors rely on organizations to create and deploy new rules as well as retest the updated WAF configuration for false positives and false negatives. However, most organizations either do not have or do not allocate sufficient time or resources to do so.

A more sustainable approach should leverage global visibility into changing threats, analyze new threats with a robust threat research capability and provide any necessary rule updates to protect against them with the least possible amount of impact on users and protected web applications.

Adequate Scale

A WAF solution without enough scale to handle the amount of incoming traffic can easily become a bottleneck, reducing web performance and possibly failing. Unfortunately, it can be difficult for organizations to predict the amount of traffic up front or quantify the scale required in a WAF solution. As a result, many organizations select a solution that does not have adequate scale and are forced to pull it out of line when the level of traffic exceeds its capabilities – either temporarily during traffic spikes or permanently as a result of web application growth – leaving the web application unprotected. A more effective approach seamlessly scales to match traffic demands as they vary over time and provide continuous protection without interruption or reducing web performance.

Ease of Management

In addition to updating for new vulnerabilities, a WAF solution needs to be continuously updated to reflect changes in the applications that it protects. This requires continuously scanning new web applications as they are first deployed as well as existing applications when they are updated, identifying new vulnerabilities and configuring rules to address those vulnerabilities. Web applications are constantly changing, and most organizations do not have the resources or expertise necessary to manage a WAF solution over time. A more manageable approach should help organizations identify rule updates that need to be made and implement them with minimal overhead.

Kona Rule Set

IBM Edge Delivery Services leverages “a proprietary rules engine called the Kona Rule Set (KRS) to employ a small number of flexible rules in conjunction with an anomaly scoring model to better address the design principles of improved accuracy and visibility into attacks.

Broader and More Flexible Rules

Rather than address every vulnerability with a dedicated rule, KRS utilizes a smaller number of broader but more flexible rules to identify malicious requests. The underlying signatures are designed for every rule to detect different attributes shared by multiple vulnerabilities, not the specific vulnerabilities themselves. This means that individual rules no longer determine if a request is malicious on their own, and KRS does not alert on or block requests based on individual rule triggers. Instead, multiple rules now work together to identify an attack.

Because every rule inspects for attributes that are common across multiple vulnerabilities, KRS has a higher likelihood of catching new attack permutations with existing rules. This improves the response that KRS provides to potential zero-day attacks – vulnerabilities that may not yet be known but have similar attributes to existing ones – and reduces the operational overhead required to manage IBM® Edge Delivery Services’ WAF solution over time.

Anomaly Scoring Model

KRS augments its WAF rules with an anomaly detection capability that provides context around individual rule triggers. Every rule trigger represents an anomaly – not a definitive conclusion, but a partial indicator that a request is malicious. In addition, different combinations of rule triggers can often be observed occurring together during different types of attacks. Creating an automated WAF testing framework enables IBM Edge Delivery Services to analyze the prevalence of every rule trigger across a wide range of known attack vectors as well as accidental by products of legitimate requests. KRS captures the observed patterns with an anomaly-scoring model.

With an anomaly-scoring model, KRS evaluates every request against the full list of enabled WAF rules and assigns a risk score based on the cumulative score of every rule triggered. KRS then alerts on or blocks a request if the cumulative risk score for that request exceeds the defined threshold for the relevant category. This provides several advantages:

- **Higher accuracy** – different rules have varying levels of accuracy in identifying malicious web requests. However, an anomaly-scoring model requires multiple rules to work together in order to determine the overall risk score of a request. This approach recognizes the role that many inaccurate rules have in helping to identify web attacks but reduces their ability to act on their own.
- **Less noise** – KRS generates an alert any time a request receives a risk score that exceeds the risk threshold, as opposed to whenever an individual rule triggers. This results in fewer alerts that administrators have to analyze and higher confidence that each alert seen represents an actual attack that must be investigated.

Weighted Risk Scoring

With an anomaly-scoring model, the scoring methodology has a significant impact on the effectiveness of the WAF solution. Different rules have varying levels of accuracy, and a well-designed rule set relies on multiple rules working together to identify an web attack. For example, one rule may be prone to false positives on its own but is indicative of an attack when triggered in conjunction with another rule. IBM Edge Delivery Services assigns every rule in KRS a weighted risk score that reflects its accuracy and contribution within the broader rule set towards identifying a malicious request. The scoring methodology for KRS relies on two IBM Edge Delivery Services capabilities:

- **Visibility** – IBM Edge Delivery Services delivers 15-30% of daily global web traffic, providing it with visibility into a substantial volume of legitimate and malicious HTTP requests targeting thousands of customer websites. This includes many known to be prone to false positives and false negatives. As a result, IBM Edge Delivery Services web security teams have a deep understanding of the characteristics demonstrated by legitimate and malicious traffic as well as scenarios that can lead to lower rule accuracy.
- **Closed-loop testing** – IBM Edge Delivery Services performs closed-loop testing to continuously measure the overall accuracy of KRS, identify sources of false positives and false negatives, and adjust score weightings as needed. Closed loop testing provides essential feedback to ensure proper weighting for every individual rule and improve overall KRS accuracy over time.

Custom Rules

In certain situations, web applications may have unique security requirements that are not covered by default with KRS. These situations can include web applications that behave abnormally or that support an organization-specific businesses process but can be easily mistaken for an web attack. IBM Edge Delivery Services' WAF solution does provide the capability to create custom rules to expand the protections of KRS and cover any unique web application and organizational requirements. However, the anomaly-scoring model combined with a mechanism to publish rule updates provides KRS with a broad, flexible and well-designed rule set that often limits the need to create custom rules.

Closed Loop Testing and Updating KRS

A web application firewall has a static configuration, protecting a defined list of web applications against a known set of threats. However, most organizations do not have sufficient resources or security expertise to continuously track developing threats, update their WAF configurations, and retest against their web traffic to ensure low false positives and false negatives. IBM Edge Delivery Services continuously updates the Kona Rule Set using closed loop testing to account for changing threats while maintaining accuracy.

Automated WAF Testing Framework

IBM Edge Delivery Services continuously tests its WAF solution using an automated WAF-testing framework. Daily test runs subject the WAF to a very large real-world set of both legitimate and malicious HTTP requests. IBM Edge Delivery Services then compares the full HTTP request and response with the expected or ideal results, analyzes root causes for false positives and false negatives, and updates existing or creates new rules. The automated WAF-testing framework provides a construct to measure WAF accuracy in a repeatable and consistent manner over time. As shown in Figure 2, this allows IBM Edge Delivery Services to better understand how changes to KRS impact overall accuracy and improve rule coverage while minimizing false positives and false negatives.

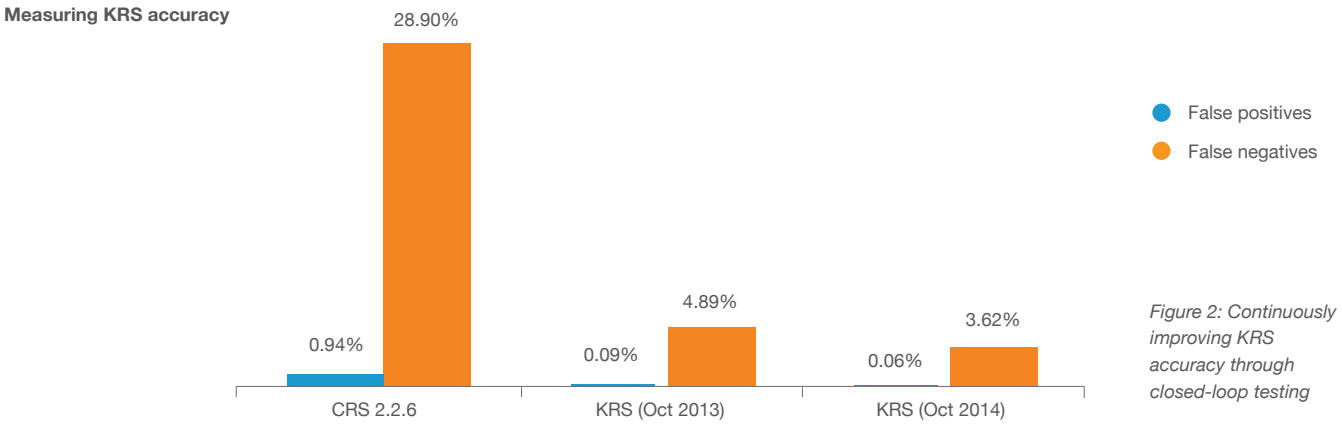


Figure 2: Continuously improving KRS accuracy through closed-loop testing

Balancing FP and FN

Figure 2 shows the accuracy of IBM Edge Delivery Services' WAF in a generic configuration before tuning for a specific customer environment. The 3.62% false-negative rate illustrates a tension present in any WAF solution – the need to balance between false positives and false negatives. When tuning KRS, IBM Edge Delivery Services' threat research team strives for the optimal balance between false positives and false negatives while working to minimize both. Examining the false negatives reveals a number of probes that pose low risk but closely resemble legitimate user requests, and further tuning KRS to identify them as malicious risks increases the false positive rate. In this case, the benefit of a lower false positive rate outweighs that of slightly higher false negatives. Organizations have the option to further tune KRS for their specific environment to remove these false negatives as well as create custom rules as necessary.

Testing with Real-world Data

Testing a WAF solution relies on a simple premise – send different attack vectors through a WAF and verify that the enabled rules stop the web attacks. However, real-world environments are more complex than test environments and often lead to false positives and false negatives. Designing a testing framework with accuracy in mind requires additional verification – not just that the tested rules detected attacks, but that they do so without inadvertently triggering false positives or false negatives. Equally important, it requires the use of real web traffic, with a large mix of both legitimate and attack traffic designed to stress the WAF response.

IBM Edge Delivery Services' automated WAF testing framework simulates live web traffic by combining a real-world set of HTTP requests with known attack vectors and exploits in a ratio of 95% legitimate to 5% attack traffic. Legitimate traffic comprises over 12,000 different HTTP requests based on recorded interactions with a large number of public websites, including:

- All of the Alexa Top 100 websites
- Websites representing a broad range of industry verticals, including e-commerce, financial services, media, social networking, and health & life sciences
- Specific websites from IBM Edge Delivery Services customers that are known to generate an above-average percentage of false positives and false negatives

Attack traffic comprises over 700 attack vectors and exploits cultivated from IBM Edge Delivery Services' Cloud Security Intelligence data analysis engine as well as publicly available tools and databases, including:

- Commercial web scanners
- Common attack tools like sqlmap and Havij
- Other known exploits from the Offensive Security Exploits Database Archive.

Publishing Rule Changes

The automated WAF testing framework allows IBM Edge Delivery Services to continuously identify rules that can be improved, create new or modify existing rules, and measure the results of any change in terms of accuracy. In addition, IBM Edge Delivery Services may release new WAF rules in response to newly discovered vulnerabilities, depending on their severity and existing coverage under available KRS rules. IBM Edge Delivery Services publishes rule changes to customers in two ways:

- **Standard** – IBM Edge Delivery Services makes regular updates to KRS as required. Rule changes are available for all customers to enable through the Luna Control Center online portal. IBM Edge Delivery Services releases new rules to KRS in situations where a large majority of customers are impacted or can benefit from the change.
- **Custom** – custom rules provide a rapid response for targeted or affected customers to implement on an individual basis. IBM Edge Delivery Services releases custom rules in situations requiring minimal implementation time or impacting a small subset of specific customers and notifies customers through their support team.

Rule Versioning

IBM Edge Delivery Services does not automatically implement rule changes for customers in order to minimize any unexpected impact on false positive and false negative. Instead, rule changes are published and customers notified of their availability through a rule-versioning feature. With rule versioning, customers can see new rules or changes to existing rules in the Luna Control Center online portal. Customers can then choose to enable individual rule changes as appropriate or necessary for their specific web application environment. Rule versioning provides IBM Edge Delivery Services customers with flexibility and granularity in configuring specific versions of individual rules within a KRS release. In addition, it allows IBM Edge Delivery Services to rapidly release new rules in response to critical vulnerabilities without the overhead of updating KRS as a whole.

Threat Intelligence

Having a robust and in-house threat intelligence capability improves a WAF vendor's ability to respond to developing threats. However, the quality, timeliness and actionability of the intelligence provided will determine the amount of impact on application security effectiveness. IBM Edge Delivery Services continuously analyzes the data available through the Akamai Intelligent Platform™ to identify current trends in the threat landscape, new attack vectors as they are first seen and currently active attackers. IBM Edge Delivery Services then incorporates that intelligence into its WAF solution in multiple ways – incident response, continuously improved WAF rules and the Client Reputation product, which allows customers to automatically block requests from IP addresses that have been rated as malicious.

Cloud Security Intelligence

Cloud Security Intelligence (CSI) provides IBM Edge Delivery Services with the mechanism to analyze attack traffic on a global scale against every customer in a timely manner. CSI utilizes Apache™ Hadoop® to ingest over 20 TB of attack data every day and retain it for 45 days, with over 2 PB of data stored at any time. CSI leverages IBM Edge Delivery Services' visibility into web traffic to thousands of the largest, most heavily trafficked and most frequently attacked online businesses to acquire relevant and high-quality data for analysis:

- **WAF rule triggers** – CSI ingests data directly from IBM Edge Delivery Services' global WAF deployments, capturing actual attack events targeting every IBM Edge Delivery Services security customer.
- **CDN logs** – CSI incorporates offline analysis performed on event logs from every IBM Edge Delivery Services customer, including those that have not deployed its WAF solution.

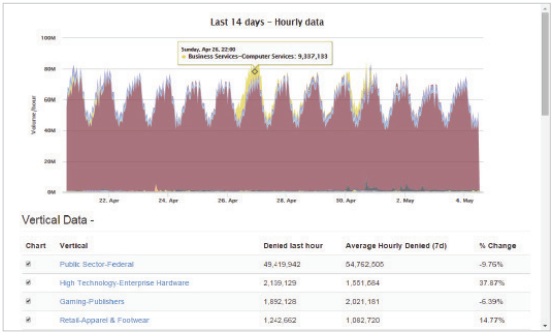


Figure 3: The IBM Edge Delivery Services WAF solution can see over 80 million rule triggers hourly across a broad range of industry verticals

Threat Research and Incident Response

Threat research and incident response organizations provide human intelligence and analysis to complement and broaden the attack coverage of a WAF solution. IBM Edge Delivery Services employs multiple teams with different charters to support its WAF customers as well as identify new attack vectors that may require additional WAF rules:

- **Threat research** – the IBM Edge Delivery Services Threat Research Team performs regular analysis of web attack trends across the entire customer base as well as custom analysis for individual customers as required. The IBM Edge Delivery Services Threat Research Team designs and implements heuristics to query CSI for actionable intelligence to support the creation of custom WAF rules, broader KRS updates and the Client Reputation product.
- **Incident response** – IBM Edge Delivery Services operates two incident response teams – the Computer Security Incident Response Team (CSIRT) and Security Emergency Response Team (SERT) – to work with IBM Edge Delivery Services' global security operations center (SOC) and provide analysis and incident response for individual customers when they experience an attack. In addition, CSIRT monitors frequently attacked IBM Edge Delivery Services customers, representing a broad range of industry verticals, as a leading indicator of new attack vectors or trends.

Client Reputation

Client Reputation augments IBM Edge Delivery Services' WAF solution with an additional layer of defense using behavioral analysis. While a WAF identifies individual malicious HTTP requests, Client Reputation identifies clients at higher risk of issuing those requests. Client Reputation performs hourly queries to CSI to identify potentially malicious clients and score them based on prior interactions with other IBM Edge Delivery Services customers. It then provides that risk score to the WAF, allowing it to alert on or block clients from issuing requests according to customizable thresholds. Client Reputation provides a simple mechanism for individual organizations to leverage IBM Edge Delivery Services' visibility into the actions of 40 million unique IP addresses on a daily basis and hundreds of millions monthly.

Case study: US national retailer
Malicious IP addresses detected over 24-hour period

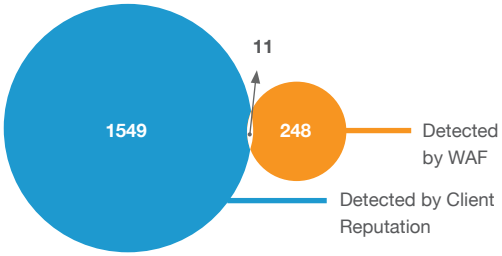


Figure 4: Comparison of malicious IP address detected by Client Reputation and the WAF over a 24-hour period

As shown in Figure 4, this visibility can help organizations detect a greater number of threats than relying on a WAF alone. Organizations can enable this protection without the overhead of managing IP whitelists and blacklists. In addition, the ability to take action based on a reputational score allows organizations to customize the protection provided to the level of risk appropriate for their business.

Globally Distributed Cloud Platform

IBM Edge Delivery Services deploys its WAF solution on a globally distributed cloud platform comprising over 189,000 servers in more than 1,400 networks and 100 countries around the world. Because users and attackers connect to protected websites through the closest IBM Edge Delivery Services server, this provides web application security at a global scale without impacting performance.

Global Scale

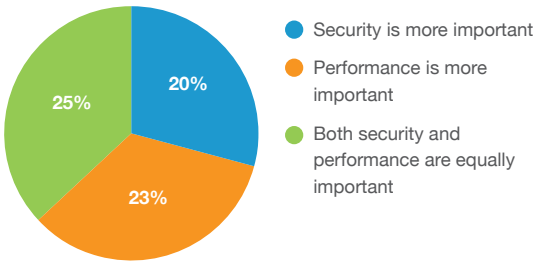
A web application firewall inspects incoming HTTP requests and evaluates the contents of each request against the list of enabled rules. For a WAF, the issue of scale revolves around both its ability to inspect the required volume of web traffic – initially and as it increases over time – and the number of WAF rules required to evaluate that traffic against. Traditional hardware-based WAF solutions often suffer from poor scale because they are limited to the CPU and memory resources available within the appliance and may have to compete with other solutions on the same appliance.

Deploying a WAF across IBM Edge Delivery Services' cloud platform eliminates the issue of scale by leveraging distributed server resources to inspect incoming web traffic. Users and attackers connect to protected websites through the closest IBM Edge Delivery Services server, which then inspects traffic for attacks and blocks any detected malicious requests. This allows IBM Edge Delivery Services' WAF solution to seamlessly scale with any increase in the amount of web application traffic – both sudden spikes in traffic as well as long-term growth – as well as with new user locations around the world.

Performance

While security solutions are not designed to improve performance, poor performance can hinder deployment of a security solution – especially a WAF solution deployed inline in front of an application. Protected websites represent critical business functions and reducing performance can lead to reduced revenue, poor user experience, or slower time to market. Figure 5 shows that IT professionals prioritize security and performance differently depending on their role in the organization. Security professionals will prioritize the quality and effectiveness of the security provided, while application owners place a premium on application performance.

Is Security or Performance more important?
Ponemon Institute, May 2015



The global scale of IBM Edge Delivery Services' cloud platform allows the WAF to protect web applications without reducing performance. The globally distributed WAF inspects HTTP traffic as it first comes onto the platform, distributing the CPU and memory resources required to inspect that traffic across all of the servers on the platform. This removes the issue of performance as a source of intra-organizational friction and an obstruction to deployment

Managed Security Services

Managing a WAF solution requires dedicated resources, with expertise in both the protected web applications as well as potential attacks. However, most organizations do not have enough staff to dedicate to managing their WAF solution or analyzing and investigating alerts. For customers deploying its WAF solution, IBM Edge Delivery Services provides two levels of managed services to help organizations monitor their protected web applications, respond to security incidents and manage their WAF over time.

Ongoing WAF Management

A web application firewall requires ongoing management in order to keep its configuration up to date with changes in protected web applications and capture newly discovered threats. IBM Edge Delivery Services helps organizations integrate this process with the natural lifecycles of their protected web applications with two capabilities:

- Regularly scheduled reviews to evaluate recent web application changes, reevaluate alert thresholds, perform false-positive and true-positive analysis, and recommend appropriate configuration updates.
- Ongoing configuration assistance to analyze proposed rule changes or available rule updates, evaluate the impact of proposed web application changes, and implement any required WAF configuration.

Managed Attack Support

In addition to ongoing WAF management, IBM Edge Delivery Services can also provide customers with managed attack support – 24/7 monitoring of protected websites and a managed response to any detected attacks. Managed attack support utilizes staff in the global SOC to respond to security incidents as they occur by:

- Responding to WAF alerts and customer requests and performing further investigation of issues.
- Determining an appropriate attack signature and deploying additional mitigation measures.
- Working with customer application teams to measure the effectiveness and accuracy of deployed mitigations, adjusting mitigations as necessary.
- Reviewing overall response with customer application teams after the incident.

Conclusion

The IBM Edge Delivery Services approach starts from an appreciation of the WAF as one of the most complex web security solutions available to organizations today. With a wide range of required security resources and expertise, few organizations have the capability to deploy and manage a WAF effectively on their own. The IBM Edge Delivery Services approach aims to make effective web application security available to any organization by simplifying much of the complexity around the WAF, as well as within the WAF itself.

The Kona Rule Set provides the foundation for IBM Edge Delivery Services' WAF solution, increasing accuracy and visibility into attacks as they occur. In addition, IBM Edge Delivery Services has constructed mechanisms around KRS to reduce the complexity of managing the WAF over time, including closed-loop testing to introduce new rules while improving accuracy, threat intelligence to keep abreast of the latest threats and managed security services to help organizations align the WAF to the lifecycles of their web applications.

Akamai® is a leading provider of cloud services for delivering, optimizing and securing online content and business applications. At the core of the company's solutions is the Akamai Intelligent Platform™ providing extensive reach, coupled with unmatched reliability, security, visibility and expertise. Akamai removes the complexities of connecting the increasingly mobile world, supporting 24/7 consumer demand, and enabling enterprises to securely leverage the cloud.