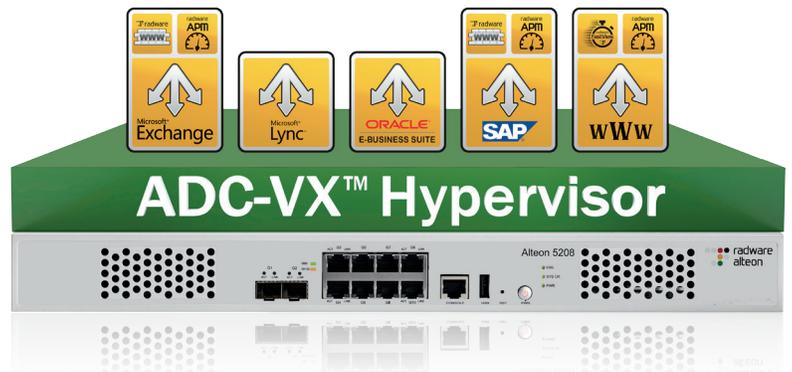




Alteon NG

The Industry's Only ADC that Delivers Full Application
SLA Assurance - Whitepaper



SHARE THIS WHITEPAPER



Table of Contents

| | |
|--|----|
| Executive Summary | 3 |
| Key IT Trends Affecting Online Application Performance | 3 |
| Virtualization, Consolidation and Shift to the Cloud..... | 3 |
| Increased Web Application Complexity | 4 |
| Mobility | 4 |
| Web Application Security..... | 5 |
| Meeting Application SLA: More Challenging than Ever | 5 |
| The Legacy ADC is Not Enough Anymore..... | 5 |
| Introducing Alteon NG: A Holistic Approach to SLA Assurance | 6 |
| Next-Generation Visibility: Application Performance Monitoring (APM) | 6 |
| Next-Generation Web Performance Optimization..... | 7 |
| Next-Generation Cyber Attack Mitigation | 8 |
| Blocking DoS Attacks Before They Block the Application | 8 |
| Application Level Protection with Advanced WAF Services | 8 |
| Summary | 10 |

Executive Summary

Emerging web application trends have created new challenges for both internal and customer facing web applications that directly affects performance. While service level agreements (SLA) for both internal and customer facing web applications have become more critical than ever, standard ADC solutions only deliver a best effort SLA and do not provide the tools required to take over the SLA of those applications.

Alteon NG, Radware’s next-generation (NG) application delivery controller (ADC), not only provides a complete set of Layer 4-7 ADC functionality, but also leverages a unique multi-service architecture that incorporates next-generation application delivery services. This enables network administrators and application/line of business (LOB) owners to proactively ensure web application SLA at all times by monitoring application performance, accelerating response time, securing the application itself and guaranteeing resources per application or service.

Key IT Trends Affecting Online Application Performance

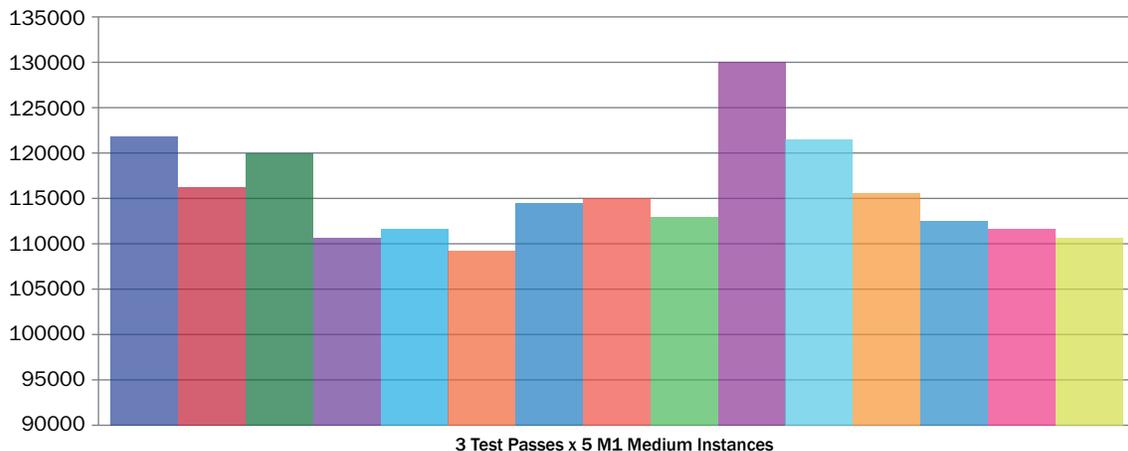
While end users have grown impatient with slow application performance, enterprise IT solutions are undergoing several transformations that negatively impact performance provided to end users.

Virtualization, Consolidation and Shift to the Cloud

Deployment of dedicated physical resources per application can guarantee the performance of applications. The utilization of these resources averages below 10% and results in cost savings through virtualization, consolidation and a cloud shared environment.

Virtualization, consolidation and cloud environments enable a higher number of applications to share a common resource pool, at a higher infrastructure utilization level that significantly reduces IT costs. In addition, leveraging datacenter infrastructure virtualization and cloud services streamlines the tedious process of rolling out a new service in-house, which requires purchasing, configuration, SW licensing and maintenance. However, this same cost model which drives the shift to the cloud also presents a new type of challenge. Applications deployed on a shared infrastructure like the cloud start competing on shared resources (especially at peak times) and deliver lower and inconsistent application performance.

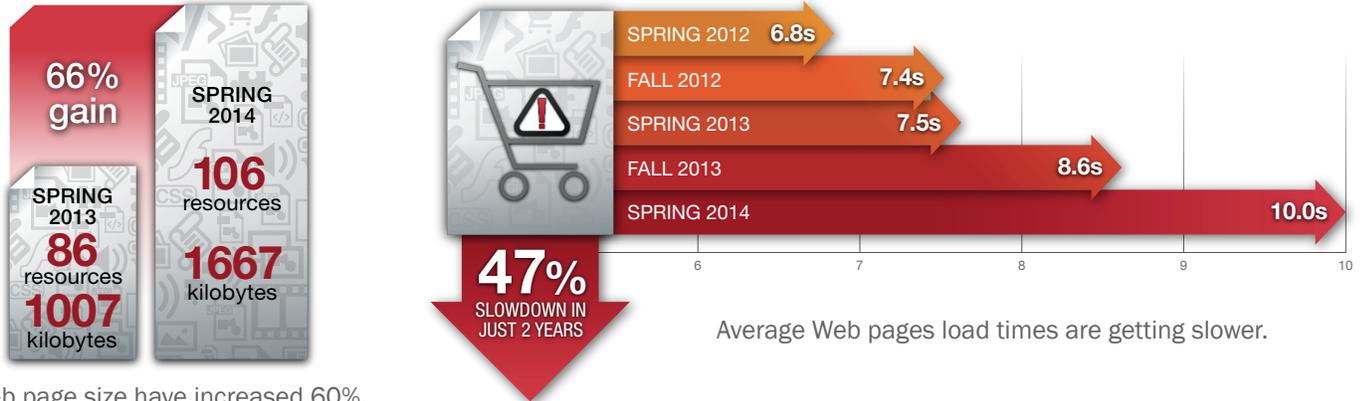
DaCapo Eclipse on M1 Medium
(milliseconds; smaller is better)



Inconsistent Cloud Service Performance
(DaCapo Benchmark Suite)

Increased Web Application Complexity

Web applications are becoming richer in functionality but result in heavier, more complex web pages with more embedded objects that impose a greater penalty on performance. For example, the average web page size grew 90% in the past two years to over 1MBytes¹. The complexity and amount of objects per web page continues to grow. It includes the number of images, cascading style sheets (CSS), and java scripts that further complicate web browser rendering. As a result, web applications are suffering a significant performance penalty that negatively affects the overall user experience.



Web page size have increased 60%,
of objects have increased 25%

Figure 1 - Web pages are getting bigger, directly affecting web application response time. (Source: State of the Union report)

Mobility

There has been a major increase in the use of applications on mobile devices. In addition, mobile networks are characterized by a higher network delay (compared to wireline networks), which results in reduced and inconsistent application performance. For example, "Each round trip can take 20-50 milliseconds for desktop computers – and up to a full second each for mobile users".²

Interestingly enough, mobile users still expect their mobile browsing experience to be faster than on a desktop computer.

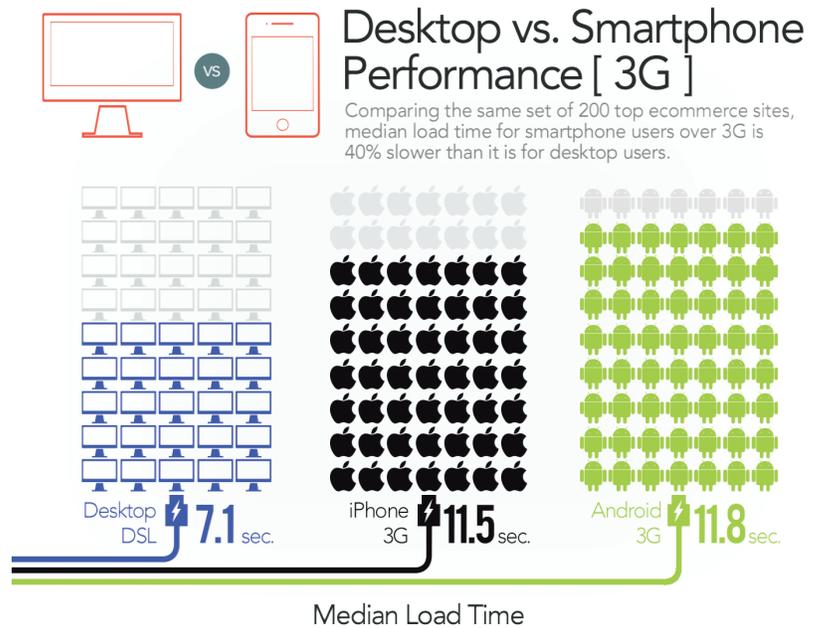


Figure 2 - Desktop computers vs. mobile device performance

¹ Source: http archive <http://archive.org/trends.php?s=All&minlabel=Sep+1+2011&maxlabel=Sep+1+2013#bytesTotal&reqTotal>

² Radware's State of the Union Report, Spring 2013

Web Application Security

More businesses are suffering from an increase in complex attacks with multiple attack vectors. In addition, attacks are lasting for longer periods of time with higher attack volumes. As a result, online applications can no longer afford to remain unprotected. However, online application protection is a resource-intensive task in terms of provisioning, maintenance and computing resources. Compromising on anything may result in security breaches and/or significant performance hits, especially when it is under attack (and when protection is needed the most).

Meeting Application SLA: More Challenging than Ever

These new trends are causing significant penalties to application performance as well as inconsistent application SLA. Unlike the past, SLA is not just about availability; it has evolved into a multi-dimension term. It's now clear that IT teams need to enforce a well-defined SLA per application, which goes beyond 24/7 availability and increased utilization. An application that is available but provides a response time of 30 seconds is not considered a well-performing application.

In addition, depending on the importance of the application to the organization's operation, each application may require a different SLA definition. For example, an online retailer may have a very strict SLA definition for their online shopping application and a lighter SLA for the employee portal.

The Legacy ADC is Not Enough Anymore

Legacy ADCs are based on a best-effort approach, where all ADC resources are shared between served applications. No isolated resources per application can impact neighboring application performance. Moreover, adding features/services to one application can degrade the overall ADC performance. In order to guarantee and enhance applications' SLA, several tools that have not traditionally been part of legacy ADC functionality are required. Legacy ADCs were never designed to guarantee applications' SLA, especially not in a multi-application/multi-tenant environment, as resources can't be locked per application and there is no fault isolation between different applications served by the same ADC.

The legacy ADC doesn't monitor the application's SLA since it doesn't gather all the relevant information (transaction completion ratio, response time etc.) nor does it have a centralized analytics/reporting engine to represent SLA information.

While web applications are suffering from performance hits, legacy ADCs still lack the ability to deliver web application performance optimization. Their ability to offload SSL and compression tasks from the web application server may optimize utilization level, but it won't improve the end user quality of experience – leaving performance acceleration uncovered.

Based on recent IT trends, a next-gen ADC is required more than ever. An ADC which is designed from the ground up to deliver a predictable, consistent high SLA with full application protection and will enable IT teams to define, monitor and actively enforce their application's SLA. These characteristics simply do not exist in current standard ADCs.

What is application SLA and how can you enforce it? A good SLA definition should include the following parameters:

1. Application's availability – the percentage of time the application must be available (e.g. 99.999%).
2. Application's performance – under various conditions like load (e.g. the amount of transactions per minute vs. response time), performance variation.
3. End user quality of experience – i.e. what is the application's response time as experienced by the end user (not just by the application admin in the datacenter).
4. Error rate – the percentage of error responses the application can deliver before breaching its SLA definition.
5. The minimum percentage of application's transactions are required to meet SLA.

Introducing Alteon NG: A Holistic Approach to SLA Assurance

Radware's Alteon next-generation (NG) range of application delivery controllers have been designed to deliver a holistic solution for SLA assurance and address the most current challenges and trends. The next generation ADC services provided by the Alteon NG allow IT administrators to gain full control over their applications' SLA and deliver an improved and consistent end user quality of experience.

Next-Generation Visibility: Application Performance Monitoring (APM)

It is impossible to manage an application's SLA without gaining visibility into it. However, old solutions for monitoring application performance and SLA has long been considered a costly and complex task which required inserting hardware probes and/or integrating software agents into every application server. Radware's Alteon NG APM service provides both network and application administrators with a simple solution, integrated into the ADC function, seamlessly providing in-depth visibility on the application's SLA.

The integration of the APM service into the Alteon NG ADC provides unique advantages. It gathers performance information from various parts of the application delivery chain; including data center performance, network performance and end user quality of experience. Coupled with an advanced centralized reporting engine, the Alteon NG APM service provides a powerful tool for real time visibility on all performance and SLA aspects. This empowers both the application administrators and the network managers to quickly detect and troubleshoot performance problems as they occur.

In the figure below, one screen snapshot provides all SLA information required including: average application/ transaction response time, responses' error rate, usage volume, performance variation and the summarizing parameter - percentage of transactions meeting SLA.

Performance issues indicated throughout this report can provide further details of the transaction level, location level (e.g. different branch offices or per country/city/ISP) and also analysis within the datacenter, network performance or end user rendering time.

Additional historical reports provide analysis of an application's infrastructure resource planning by analyzing performance results vs. amount of transactions. Using the performance figure breakdown of delay contribution of datacenters, network and user rendering - IT administrators can easily understand if application resources are enough or not; analyzing whether the network resources are the bottleneck or compute resources, etc.

It's easy to gain visibility into an applications SLA, by activating the APM service on the Alteon NG. It works right out of the box and does not require any integration or change in the application server or code, nor does it require any hardware probes or synthetic transactions to provide this real time, real life performance information. In case of changes to the application, the Alteon NG APM service automatically detects the new type of transaction and adds performance information to its reports, keeping even the APM service maintenance simple and seamless.

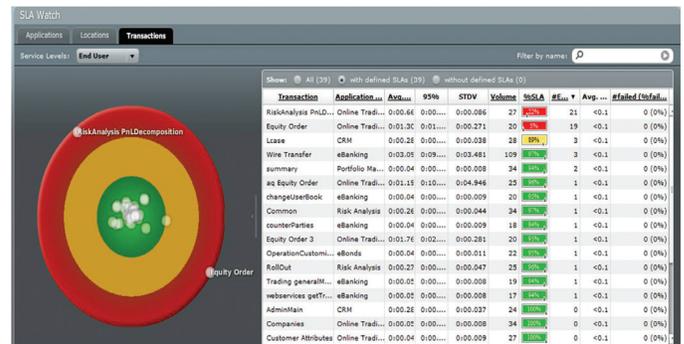


Figure 3 - Alteon NG detailed application performance report



Figure 4 - Historical performance report with detailed breakdown per application delivery segment.

Next-Generation Web Performance Optimization

Alteon NG significantly improves application acceleration with a new web performance optimization (WPO) capability that decreases web page build times by up to 50% on both PC and mobile browsers.

FastView, employs industry leading technology to optimize the code of actual web pages so they render more quickly in a client browser. With web pages increasing in size and complexity, latency and render times have become significant sources of delay, especially in a mobile environment. FastView uses a variety of techniques to rewrite HTML and combine web objects. This allows pages to automatically be compiled and optimized for different desktop and mobile browsers – saving on manual optimization efforts and QA costs. The simplicity of the FastView deployment and its seamless adaptation to different web applications and application changes allows to offload web application code optimization work from the programmers team to an expert device and focus on core business competencies.

Radware's FastView acceleration technology employs 22 separate acceleration treatments for different application and browser scenarios, including:

- **Simplifying large, complex web pages** – the average web page is already over 1.5MByte and contains 100 objects. FastView acceleration technology automatically consolidates similar objects, such as images, CSSs, Java Scripts, and etc. This reduces the number of browser requests per page and the number of round trip delays, making each page load as fast as possible.
- **Caching** – smart caching algorithm in FastView service ensures end user clients will never download the same object twice from the server, while guaranteeing the end user will always receive the most up-to-date content. The result is faster application response time, offloading browser-server requests and corresponding server processing.
- **Acceleration for the entire web transaction not just single web pages** – FastView learns and predicts where online visitors are likely to go next. It preloads the relevant elements of subsequent web pages into the browser's cache to have them locally on standby resulting in faster response time for multi-page transaction.
- **Third-Party timing and SLAs** – retrieve third-party content as well as create third-party service level agreements (SLA) that allot a maximum wait time for each third-party script on a page. If the script has not loaded within this time frame, it is deferred until after the remainder of the page renders and/or is cancelled altogether.
- **Recognizing that not all browsers are created equal** – retrieve third-party content as well as create third-party service level agreements (SLA) that allot a maximum wait time for each third-party script on a page. If the script has not loaded within this time frame, it is deferred until after the remainder of the page renders and/or is cancelled altogether.
- **Content Minification** – reduce content size by removing and trimming redundant data from web pages.
- **Acceleration for mobile devices:**
 - **Mobile Caching** – mobile devices often use a very limited and useless cache mechanism. FastView automatically creates a dedicated caching solution for the web application on any mobile browser, based on HTML5 local storage.
 - **Image resizing** – FastView can automatically detect requests for images from devices with smaller screens and automatically resize images according to the device's size, reducing the file size and simplifying image rendering.
 - **Touch-to-click conversion** – while mobile devices use touch screens, any finger press on a link needs to be converted to a mouse click by the web application – taking up to half a second of delay. FastView eliminates this delay by automatically converting touches to clicks for mobile clients.

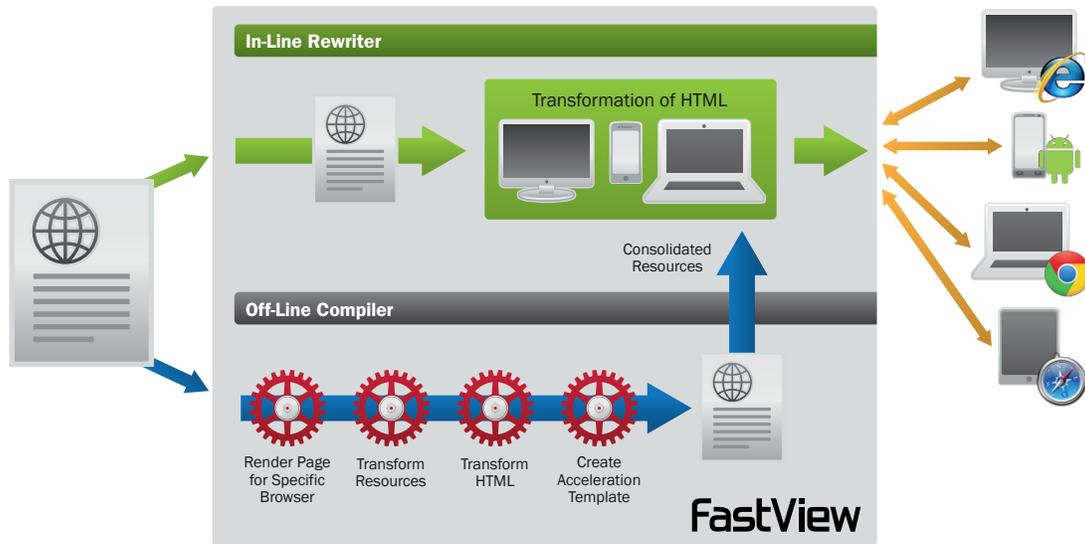


Figure 5 - FastView offline learning engine for real time web performance optimization per browser.

Next-Generation Cyber Attack Mitigation

Securing the application’s infrastructure requires a layered approach that protects all segments that delivers the application. As an integral part of the organizational security architecture, Alteon NG includes several next generation security services that fully integrates into [Radware’s Attack Mitigation Solution \(AMS\)](#), adding another important layer to protect the application it serves.

Blocking DoS Attacks Before They Block the Application

The Alteon NG ADC is situated in a strategic location of the application delivery chain, overlooking all application related traffic. By inspecting various traffic parameters (e.g. bandwidth, PPS, CPS, CEC) as well as ADC health parameters (e.g. CPU utilization, network table’s capacity) under normal conditions, it can set a base line for normal traffic and effectively identify anomalies related to cyber attacks. The Alteon NG can mitigate those attacks by signaling the other components of the AMS solution, blocking the attack further away from the application itself.

- Signaling to DefensePro mitigates the attack at the perimeter of the organization’s network, much before it even enters it.
- Signaling to DefensePipe pushes volumetric attacks to a scrubbing center in the cloud, eliminating risks of saturating the internet connection pipe of the organization.

As a result attacks can be detected anywhere (i.e. at the edge of the network, inside the datacenter etc.) and mitigated in the best location, to ensure consistent application SLA even under attack.

Application Level Protection with Advanced WAF Services

Advanced load balancing Web Application Firewall (WAF) services provide a true multi-application architecture that is fault tolerant and scalable. Legacy ADC’s integrated WAF, together with other Layer 4-7 services on the same ADC device, are consuming the same shared computing resources that result in overall performance degradation and potential SLA breaches. In contrast, Alteon NG leverages a true multi-application architecture that integrates Radware’s AppWall - a WAF module, that has its own dedicated and pre-allocated resources. This is in addition to the vADC resources it serves – ensuring no impact on neighboring applications SLA or neighboring ADC services.

Alteon NG integrates the AppWall service that provides both application level attack detection as well as mitigation that blocks the attack in the device. Moreover, thanks to the signaling mechanism mentioned above, the Alteon NG’s

AppWall service can also send Layer 7 attack traffic signatures to the DefensePro and/or DefensePipe. As a result, the mitigation function is pushed to the edge of the network or to the cloud, before it enters the datacenter’s network and risks additional devices (e.g. firewalls, routers ADCs etc.) in the application delivery chain.

Alteon NGs integrated AppWall service provides a patent-protected technology to create and maintain security policies for the widest security coverage, with the lowest false positives and minimal operational effort. Once enabled, the auto policy generation module within the AppWall service module, analyzes the security related attributes of the protected web application and derives the potential threats in the application. The web application is mapped into application zones, each with its own common potential threats. Finally it generates individual, granular protection rules for each zone and sets a policy in blocking mode. Once it has completed the optimization process, false-positives are minimized and maintain best security coverage.

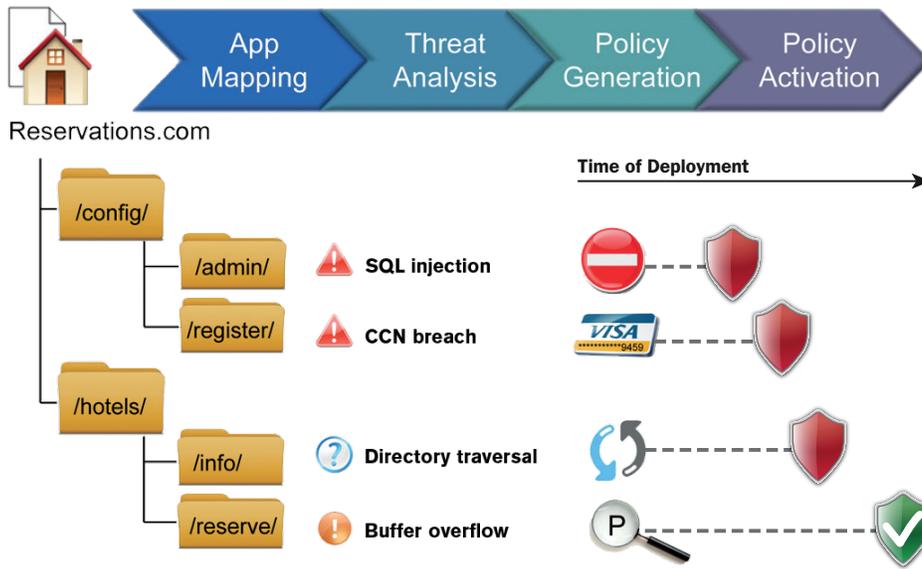


Figure 6 - AppWall's auto learning process

There are several benefits to the Alteon NG security service. While most WAF services require long and tedious manual configuration with long QA test cycles, the Alteon NGs AppWall service provides an automated and reliable learning and provisioning technology which minimizes deployment efforts without compromising on security or application SLA.

Coupled with APSolute Vision, Radware’s centralized management application; customers gain end-to-end visibility through a single pane of glass, for both the ADC services and the AMS services, including the Alteon NGs AppWall service. Vision provides a centralized device and service management as well as in-depth reporting and security event management for full attack visibility and forensic.

Next-Generation, Multi-Service Architecture for Full SLA Assurance

Alteon NG provides the industry’s first fault isolated instance that provide virtualization while locking in machine resources per virtual instance for enterprises at any size. This technology enables each Alteon NG platform to run multiple, completely autonomous ADC instances, each with its independent operating system versions, CPU cores, memory, network stack and management control.

The ability to create ADC instances and allocate dedicated system resources to each and within the Alteon NG platform means that SLA requirements for each ADC instance can always be met and web advanced services can run without impacting performance.

Legacy ADC approaches with shared virtual segmentation result in resource contention between ADC instances – where one overloaded application will hog system resources from another. Organizations that need to ensure the performance of critical applications can dedicate a virtual ADC instance to each application and be assured that the performance of the application will never be compromised. The unique ADC deployment model in Alteon NG provides fully fault isolated ADC instances for running multiple services.

The Alteon NG architecture was designed to dynamically scale when necessary. On demand scalability provides more throughput, services and vADCs into existing Alteon platforms, with no hardware modifications and without causing any disturbance to running applications or risking resource contention on any of the ADC services. Each Alteon NG instance can be allocated with capacity units as needed.

Moreover, next generation ADC services are often resource intensive and their resource requirement may vary significantly in different scenarios (e.g. under cyber attack, or serving more complex pages) and across applications. Radware's Alteon NG solution offers a deployment model which not only locks resources per application, but also guarantees resources per NG ADC service such as FastView and WAF, allocating dedicated capacity units per service.

Should additional resources be required, Alteon NG also supports OnDemand scaling out of the box, leveraging external compute resources for those computational intensive NG services.

Summary

Emerging trends in the IT and application infrastructure require a different approach to delivering applications and guaranteeing SLA. Unlike legacy ADC solutions which only deliver best effort application SLA, the Alteon NG solution is designed from the ground up to provide a holistic solution for SLA assurance and SLA improvement, enabling proactive application performance management. The Alteon NG features a unique architecture that guarantees resource reservation and SLA per application and per next generation ADC service. It allows simple scalability and resource allocation and next-generation services that extend to in-depth visibility into an application's SLA with APM. Additionally, it provides tools to improve SLA with FastView advanced performance acceleration service and simplicity in the deployment of automated Alteon NG's WAF security services that cannot degrade neighboring applications SLA. And finally, integration into Radware's AMS that pushes attacks away from the datacenter. Unlike any other ADC, the combination of Alteon NG services and its architecture empowers IT administrators to take control and guarantee application SLA.

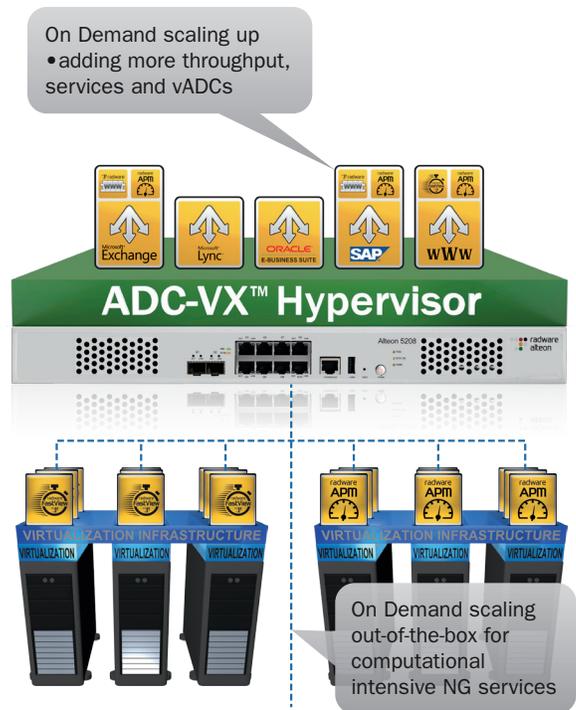


Figure 7 - Multi service architecture for full SLA assurance