# The 2012
# Application & Service Delivery Handbook

Part 4: Planning, Management and Security

***By Dr. Jim Metzler, Ashton Metzler & Associates***
***Distinguished Research Fellow and Co-Founder***
***Webtorials Analyst Division***

Platinum Sponsors:

Gold Sponsors:

Produced by:

# Planning, Management and Security

# Executive Summary

The ***2012 Application and Service Delivery Handbook*** will be published both in its entirety and in a serial fashion. This is the fourth of the serial publications. The first publication focused on describing a set of factors, such as chatty protocols, that have traditionally complicated the task of ensuring acceptable application delivery. The second publication described a set of emerging challenges, such as the movement to bring your own device to work, that are beginning to impact the ability of IT organizations to ensure acceptable application and service delivery. The third publication described the technologies and services that are available to improve the performance of applications and services.

The primary goal of this publication is to describe the technologies and services that are available to improve the management and security of applications and services. The fifth and final publication will include an executive summary as well as a copy of the complete document.

A preceding section of The ***2012 Application and Service Delivery Handbook*** described the surveys that were administered to the subscribers of Webtorials. Throughout this document, the IT professionals that responded to those surveys will be referred to as The Survey Respondents.

# Planning

## Background

In the classic novel *Alice in Wonderland,* English mathematician Lewis Carroll first explained part of the need for why planning is important to application and service delivery (though he may not have known it at the time). In the novel, Alice asks the Cheshire cat, "Which way should I go?" The cat replies, "Where do you want to get to?" Alice responds, "I don't know," to which the cat says, "Then it doesn't much matter which way you go."

***Hope is not a strategy. Successful application and service delivery requires careful planning.***

Many planning functions are critical to the success of application delivery. One planning function that has been previously discussed, and will be discussed again in a subsequent sub-section of this handbook, is identifying the company's key applications and services and establishing SLAs for them. As described in that sub-section, it is not sufficient to just establish SLAs for the company's key applications and services. IT organizations must also identify the key elements (e.g., specific switches and routers, WAN links, servers, virtual machines) that support each of the applications. Other key steps include:

- Baselining the performance of each of the organization's critical applications.
- Baselining the performance of each of the key elements that support each of the critical applications and identifying at what levels of utilization and delay the performance of each of the elements has an unacceptable impact on the performance of the application.
- Establishing SLAs for each of the key elements.

Another key planning activity that is discussed in a subsequent sub-section of this handbook is Application Performance Engineering (APE).

***The primary goal of APE is to help IT organizations reduce risk and build better relationships with the company's business unit managers.***

APE achieves this goal by anticipating and, wherever possible, eliminating performance problems at every stage of the application lifecycle.

Another key planning activity is performing a pre-deployment assessment of the current environment to identify any potential problems that might affect an IT organization's ability to deploy a new application. One task that is associated with this activity is to either create or update the IT organization's inventory of the applications running on the network. Part of the value of this task is to identify unauthorized use of the network; i.e., on-line gaming and streaming radio or video. Blocking unauthorized use of the network can free up additional WAN bandwidth. Another part of the value of this task is to identify business activities, such as downloads of server patches that are being performed during peak times. Moving these activities to an off-peak time also releases additional bandwidth.

Another task associated with performing a pre-deployment assessment is to create a current baseline of the network and the key applications. Relative to baselining the network, IT

organizations should modify how they think about baselining to focus not just on utilization, but also on delay. In some instances, however, even measuring delay is not enough. If, for example, a company is about to deploy an application such as telepresence then the pre-assessment baseline must also measure the current levels of jitter and packet loss. Relative to baselining the company's key applications, this activity involves measuring the average and peak application response times for key applications, both before and after the new application is deployed. This information will allow IT organizations to determine if deploying the new application caused an unacceptable impact on the company's other key applications.

## Integrating Network Planning and Network Operations

As noted, the next section of the handbook discusses APE. One of the characteristics of APE is that it is a life cycle approach to planning and managing application performance. Addressing performance issues throughout the application lifecycle is greatly simplified if there are tight linkages between the IT personnel responsible for the planning and operational functions. The degree of integration between planning and operations can be significantly enhanced by a common tool set that:

- Provides estimates of the impact on both network and application performance that would result from proposed changes in either the infrastructure or in application traffic patterns.

- Verifies and ensures consistency of configuration changes to ensure error-free network operations and satisfactory levels of service

A common tool set that spans planning and operational functions also supports initiatives aimed at the consolidation of network management tools the goal of which is to reduce complexity and maximize the productivity of the IT staff.

For those organizations that run a large, complex network there often is a significant gap between network planning and network operations. One of the reasons for this gap is that due to the complex nature of the network there tends to be a high degree of specialization amongst the members of the IT function. Put simply, the members of the organization who do planning understand planning, but typically do not understand operations. Conversely, the members of the organization who do operations understand operations, but typically do not understand planning.

Another reason for this gap is that historically it has been very difficult to integrate planning into the ongoing change management processes. For example, many IT organizations use a change management solution to validate changes before they are implemented. These solutions are valuable because they identify syntax errors that could lead to an outage. These solutions, however, cannot identify how the intended changes would impact the overall performance of the network.

## Route Analytics

A class of management tool that can facilitate the integration of planning and operations is typified by an IP route analytics solution[1].

***The goal of route analytics is to provide visibility, analysis and diagnosis of the issues that occur at the routing layer in complex, meshed networks.***

A route analytics appliance draws its primary data directly from the network in real time by participating in the IP routing protocol exchanges. This allows the route analytics solution to compute a real-time Layer 3 topology of the end-end network, detect routing events in real time and correlate routing events or topology changes with other information, including application performance metrics. As a result, route analytics can help determine the impact on performance of both planned and actual changes in the Layer 3 network.

Route analytics is gaining in popularity because the only alternative for resolving logical issues involves a very time-consuming investigation of the configuration and log files of numerous individual devices. As described in the next section of the handbook, a logical issue such as route flapping typically causes notably more business disruption than does a physical issue and a logical issue typically takes notably longer to troubleshoot and repair than does a physical issue.

Route analytics is also valuable because it can be used to eliminate problems stemming from human errors in a router's configuration by allowing the effect of a configuration change to be previewed before the change is actually implemented. From an application delivery perspective, route analytics allows the path that application traffic takes through the network to be predetermined before changes are implemented and then allows the application traffic to be tracked in real-time after the application has gone into production.

## Planning for Cloud Computing

Most IT organizations that have already implemented either public or private cloud computing have not done so in a highly systematic fashion. In some cases, they used a trial and error approach to choosing a SaaS provider, while in other cases they evaluated one aspect of private cloud computing (e.g., server virtualization) without considering other aspects of private cloud computing and did not plan for the impact that server virtualization would have on other components of IT, such as management or the design of the data center LAN.

***In order to maximize the benefit of cloud computing, IT organizations need to develop a plan (The Cloud Computing Plan) that they update on a regular basis.***

The Cloud Computing Plan should identify the opportunities and risks associated with both public and private cloud computing. The Cloud Computing Plan must identify a roadmap of what steps the IT organization will take on a quarter-by-quarter basis for the next two to three years and ensure that the steps are in line with the corporate culture. This includes identifying:

[1] More information on this topic can be found at: Webtorials.com

- What functionality (e.g., applications, storage) needs to remain under the tight control of the IT organization and what functionality is appropriate to hand over to a Cloud Computing Service Provider (CCSP).

- What levels of service are good enough for each class of application and for the myriad storage and compute requirements.

- How the IT organization will evolve over time the twelve characteristics of a cloud computing solution that were discussed in a previous section of the handbook; e.g., virtualization, automation, simplification.

- How the IT organization will evolve its data center LAN architecture to support private cloud computing.

- How the IT organization will evolve its use of WAN services to support all forms of cloud computing.

- How the IT organization will minimize the security and confidentiality risks associated primarily with public cloud computing services, but also with private cloud computing.

- What management functionality must be present in the management domain controlled by the IT organization as well as provided by the relevant network service providers and CCSP(s).

- How the IT organization will overcome potential performance bottlenecks.

The Cloud Computing Plan should look systematically across multiple technologies because of the interconnected nature of the technologies. As part of creating this plan, IT organizations need to understand the cloud computing strategy of their existing and potential suppliers, including the partnerships that the suppliers are establishing between and amongst themselves.

# Management

## Background

As will be discussed in this section of the handbook, in order to respond to the myriad challenges facing them, IT organizations need to adopt an approach to service management that is comprised of the following four components:

- Either a multi-tier application or multiple applications
- Supporting protocols
- Enabling network services; e.g., DNS, DHCP
- The end-to-end network

## Limitations of the Traditional Approaches to Management

The almost total reliance that organizations have on networked applications and services has pushed the task of ensuring acceptable application and service performance up to the critical level for almost all organizations. Despite this high priority, the traditional network and application performance management systems have lagged behind.

**Traditional Network Performance Management Systems**

Most Network Performance Management Systems (NPMS) had their origins in monitoring the performance of telecommunication carriers to verify that organizations were getting the services they paid for. These systems are based on a combination of Simple Network Management Protocol (SNMP) and Internet Control Message Protocol (ICMP, also known as "ping"). Traditional NPMS measured how long it took a packet to travel from the data center to the branch office network and back - thus determining the Round Trip Time (RTT). If the return packet did not arrive within a few seconds, the original packet was deemed lost and this is how packet loss was measured.

These early NPMS solution worked acceptably well for traditional client/server applications and other centrally hosted applications. However, as technology and applications evolved, the limitations of these systems became apparent. Those limitations include the fact that early NPMS systems:

- Only measured from the central data center to the edge of the branch office network. Problems inside the branch office network went unreported until end users complained.
- Had difficulty measuring network paths outside of the data center, such as those used by VoIP, IP Video and other peer-to-peer communication traffic.
- Measured performance across the entire path but did not isolate which network segments have performance issues.

## Traditional Application Performance Management

As application architectures evolved from client/server to n-tier web-based applications, application functionality on the server was usually divided up into two or three segments. These segments are the web front-end (presentation tier or tier 1), business logic processes (logic tier or tier 2) and database operations (data tier or tier 3).

In an n-tier web based application, the user interacts with the presentation tier and the presentation tier in turn communicates to the logic tier, which in turn communicates to the data tier. Each tier uses servers that are optimized to the characteristics of their tier. A presentation tier server, for example, is optimized for network I/O and web traffic; e.g. multiple network cards, large network buffers, etc. A logic tier server is optimized for logic computations; e.g. high speed CPUs, large memory size, etc. A data tier server is optimized for database operations; e.g. multiple disk I/O controllers, large disk cache, large memory size, etc.)

Traditional application performance management (APM) was typically performed separately from network performance management. For example, when application degradation occurs, the triage process typically assigns the incident to either the network or server areas for resolution. Each area then examines their basic internal measurements of network and server performance and a pronouncement is made that the source of the issue is either the network or the application server or both or neither. Since these tasks are typically done by different parts of the IT organization using different toolsets and management frameworks, it is quite possible that conflicting answers are given for the source of application performance issues.

Similar to traditional NPMS, traditional APM solutions have limitations. Those limitations include the fact that that traditional APM solutions:

- Only describe the performance within a single server, not the combined performance across all tiers of an application.
- Cannot attribute CPU, disk I/O, network I/O nor memory utilization to specific classes of transactions. Only aggregate server performance information is available.
- Do not integrate network performance data between tiers to monitor and analyze application performance problems.

## Synthetic Transactions

Synthetic transactions provide a somewhat more realistic measurement of application performance than traditional NPMS and APM solutions. While synthetic transactions have the advantage of being a better representation of the end user's experience, they also have several disadvantages, including:

- The application being monitored has to be constructed to allow transactions that have no business impact. For example, a banking application would have to have a special account so that when money was added or subtracted from this special account, it would not count towards the banks total assets.

- Synthetic transactions frequently originate from the same data center in which the application servers reside and are not subject to the typical network latencies and availabilities that are present in branch office networks.

- Frequently exercising a synthetic transaction can cause the transaction to perform notably differently than a real production transaction would. For example, a frequently exercised transaction may have its related data in cache all the time and not loaded from disk. As a result, the synthetic transaction would occur notably quicker than a production transaction would.

## Forces Driving Change

Previous sections of this handbook detailed the traditional and emerging service and application delivery challenges. This subsection will identify how some of those challenges are forcing a change in terms of how IT organizations perform services.

### Server Virtualization

Until recently, IT management was based on the assumption the IT organizations performed tasks such as monitoring, baselining and troubleshooting on a server-by-server basis. Now, given the widespread adoption of server virtualization, the traditional approach to IT management must change to enable management tasks to be performed on a virtual machine (VM)-by-VM basis. Another assumption that underpinned the traditional approach to IT management was that the data center environment was static. For example, it was commonly assumed that an application resided on a given server, or set of servers, for very long periods of time. However, part of the value proposition that is associated with server virtualization is that it is possible to migrate VMs between physical servers, both within the same data center and between disparate data centers.

***IT organizations need to adopt an approach to management that is based on the assumption that the components of a service, and the location of those components, can and will change frequently.***

### Cloud Balancing

IT management has historically been based on the assumption that users of an application accessed that application in one of the enterprise's data centers and that the location of that data center changed very infrequently over time. The adoption of IaaS solutions in general, and the adoption of cloud balancing in particular demonstrates why IT organizations need to adopt an approach to IT management that is based on gathering management data across myriad data centers, including ones that are owned and operated by a third party. The adoption of cloud balancing is also another example of why IT organizations need to adopt an approach to management that is based on the assumption that the components of a service, and the location of those components, can and will change frequently.

### Delay Sensitive Traffic

Voice and video are examples of applications that have high visibility and which are very sensitive to transmission impairments. As part of the traditional approach to IT management it is common practice to use network performance measurements such as delay, jitter and packet

loss as a surrogate for the performance of applications and services. A more effective approach is to focus on aspects of the communications that are more closely aligned with ensuring acceptable application and service delivery. This includes looking at the application payload and measuring the quality of the voice and video communications. In the case of UC, it also means monitoring the signaling between the components of the UC solution.

In addition to having a single set of tools and more of a focus on application payload, IT organizations need to implement management processes that understand the impact that each application is having on the other applications and that can:

- Analyze voice, video, UC and data applications in consort with the network
- Support multi-vendor environments
- Support multiple locations

**Converged Management**

As mentioned in the section of the report that focused on the emerging application and service delivery challenges, one of the characteristics of cloud computing is the integration of networking, servers and computing in the data center. While a converged data center infrastructure offers a number of benefits, it does create a number of management challenges. These challenges generally fall into the following categories:

- Implementing control of the infrastructure in a significantly more efficient manner than is possible with the traditional labor-intensive tool sets that are not well integrated or automated.
- Ensuring compliance with enterprise policies, as well as best-practice guidelines for service deployment in the converged infrastructure.
- Implementing granular, near real-time provisioning and change management for virtualized infrastructure services.
- Expediting resolution of any issues that could compromise the performance or availability of infrastructure services.

Meeting these challenges will involve a number of changes in the way the data center is managed. In particular, the converged infrastructure requires a management system and management processes that have the same level of integration and cross-domain convergence that the infrastructure has. For example, in order to support the requirement for the dynamic provisioning and re-allocation of resources to support a given IT service, the traditional manual processes for synchronizing the required server, network and storage resources will have to be replaced with integrated, automated processes. In order to enable this change, the provisioning and change management processes will need to be integrated and will need to feature the automatic configuration of network and storage resources when additional infrastructure services are deployed or when additional physical or virtual servers are brought on line or are moved. In a similar fashion, operations management needs to be consolidated and automated to keep service quality in line with user expectations.

IT departments can take a do-it-yourself (DIY) approach to implementing an integrated, cross-domain management system. They could, for example, leverage the available element manager plug-ins and APIs and piece together a management system that integrates at least some of the overall components of the infrastructure. However, this type of ad hoc automation and integration across the end-to-end, cross-domain infrastructure is quite time-consuming and involves considerable specialized programming expertise and detailed technical knowledge. It is also expensive and time consuming to support over time. There is an alternative to the DIY approach. IT departments that are evaluating converged infrastructure solutions can expect to see highly effective pre-packaged integrated management solutions that negate the need for a DIY approach. In addition, IT organizations can use the breadth, efficiency, and effectiveness of the converged management solutions as one of the key decision criteria to evaluate the converged infrastructure solutions that are offered by competing vendors.

While not a requirement, the cross-domain integrated management of a converged infrastructure will bring the greatest benefit in those instances in which a single administrator has the authority to initiate and complete cross-domain management tasks, such as provisioning and modifying infrastructure services. For example, the use of a single administrator can eliminate the considerable delays that typically occur in a traditional management environment where the originating administrator must request other administrators to synchronize the configuration of elements within their domains of responsibility. However, in many cases the evolution from the current approach of having separate administrators for each technology domain to an approach in which there is a single administrator will involve organizational challenges. As a result, many IT organizations will evolve to this new approach slowly over time.

## Mobile Device Management

In the current environment, it is possible to find a Cloud-based service that provides almost any possible form of management functionality. For example, there are multiple Cloud-based services currently available in the marketplace that manage network equipment such as routers, firewalls and IPSs.

The section on the handbook that described the first generation of application and service delivery challenges discussed the fact that the IBM X-Force 2011 Trend and Risk Report[2] highlighted the fact that new trends such as the ongoing adoption of mobile devices creates challenges for enterprise management and security. For example, the IBM stated that in 2011 there was a 19 percent increase over 2010 in the number of exploits publicly released that can be used to target mobile devices such as those that are associated with the movement to Bring your Own Device (BYOD) to work. The report added that there are many mobile devices in consumers' hands that have unpatched vulnerabilities to publicly released exploits, creating an opportunity for attackers.

One way to respond to the challenges associated with mobile devices is by using a Cloud-based mobile device management service. In order to be an effective, such a service should:

- Offer a portal that provides a comprehensive view into the mobile device environment.
- Given the ongoing consumerization of IT, the portal should provide visibility into both corporate and employee-owned mobile devices.

[2] X-Force 2011 Trend and Risk Report

- Enable the organization to implement management security policies through simple click configuration and provisioning.

- Control access to key network resources and applications by specific user groups or by lines of business.

- Detect if a mobile device has been lost and if so, delete the data from the device.

- Detect when roaming, data, voice, or SMS usage thresholds have been reached, and provide real-time alerts to prevent unwanted billing overages.

## Evaluating Cloud Computing Service Providers

When It organizations are evaluating the adoption of public cloud computing solutions, they need to evaluate the CCSP's management capabilities. This sub-section contains two sets of criteria that can be used to evaluate those capabilities. The first set is focused on traditional management functionality and the second set focuses on managing the performance of applications and services.

**Traditional Management Functionality**

- What is the ability of the CCSP to manage the challenges associated with virtualization that were previously discussed in this handbook?

- What management data will the CCSP make available to the IT organization?

- What is the ability of the CCSP to troubleshoot performance or availability issues?

- What are the CCSP's management methodologies for key tasks such as troubleshooting?

- Does the CCSP provide tools such as dashboards to allow the IT organization to understand how well the service they are acquiring is performing?

- Does the CCSP provide detailed information that enables the IT organization to report on their compliance with myriad regulations?

- What are the primary management tools that the CCSP utilizes?

- What is the level of training and certification of the CCSP's management personnel?

- What are the CCSP's backup and disaster recovery capabilities?

- What approach does the CCSP take to patch management?

- What are the specific mechanisms that the IT organization can use to retrieve its data back in general and in particular if there is a dispute, the contract has expired or the CCSP goes out of business?

- Will the IT organization get its data back in the same format that it was in when it was provided to the CCSP?
- Will the CCSP allow the IT organization to test the data retrieval mechanisms on a regular basis?
- What is the escalation process to be followed when there are issues to be resolved?
- How can the service provided by the CCSP be integrated from a management perspective with other services provided by either another CCSP and/or by the IT organization?
- How can the management processes performed by the CCSP be integrated into the end-to-end management processes performed by the IT organization?

**Managing Application and Service Performance**

- What optimization techniques has the CCSP implemented?
- What ADCs and WOCs does the CCSP support?
- Does the CCSP allow a customer to incorporate their own WOC and/or ADC as part of the service provided by the CCSP?
- What is the ability of the CCSP to identify and eliminate performance issues?
- What are the procedures by which the IT organization and the CCSPs will work together to identify and resolve performance problems?
- What is the actual performance of the service and how does that vary by time of day, day of week and week of the quarter?
- Does the IT organization have any control over the performance of the service?
- What technologies does the CCSP have in place to ensure acceptable performance for the services it provides?
- Does the CCSP provide a meaningful SLA? Does that SLA have a goal for availability? Performance? Is there a significant penalty if these goals are not met? Is there a significant penalty if there is a data breach?
- To what degree is it possible to customize an SLA?
- What is the ability of the CCSP to support peak usage?
- Can the CCSP meet state and federal compliance regulations for data availability to which the IT organization is subject?

# Application Performance Management

## Background

This section of the handbook will outline an approach that IT organizations can utilize to better manage application and service delivery, where the term *service* was previously defined. However, in an effort to not add any more confusion to an already complex topic, instead of using a somewhat new phrase *application and service delivery management*, this section will use the more commonly used phrase *application performance management (APM)*.

APM is a relatively new management discipline. In spite of the newness of APM, over a quarter of The Survey Respondents said that APM was currently important to their organization and another one third of The Survey Respondents said that it is important to their organization to get better at APM. In addition to the fact that APM in general is important to IT organizations, some specific components of APM are particularly important. For example, as described below, a critical component of APM is the adoption of service level agreements (SLAs). As described in a preceding section of the handbook, over half of The Survey Respondents indicated that over the next year it is either very important or extremely important for their organization to get better at managing SLAs for one or more business critical applications.

Since any component of a complex service can cause service degradation or a service outage, IT organizations need a single unified view of all of the components that support a service. This includes the highly visible service components such as servers, storage, switches and routers, in both their traditional stand-alone format as well as in their emerging converged format; i.e., Cisco's UCS and VCE's Vblock platforms. It also includes the somewhat less visible network services such as DNS and DHCP, which are significant contributors to application degradation. Multiple organizational units within the IT organization have traditionally provided all of these service components. On an increasing basis, however, one or more network service providers and one or more cloud computing service providers will provide some or all of these service components and so in order to achieve effective service delivery management, management data must be gathered from the enterprise, one or more Network Service Providers (NSPs) and one or more CCSPs. In addition, in order to help relate the IT function with the business functions, IT organizations need to be able to understand the key performance indicators (KPIs) for critical business processes such as supply chain management and relate these business level KPIs to the performance of the IT services that support the business processes.

IT organizations must also be able to provide a common and consistent view of both the network and the applications that ride on the network to get to a service-oriented perspective. The level of granularity provided needs to vary based on the requirements of the person viewing the performance of the service or the network. For example, a business unit manager typically wants a view of a service than is different than the view wanted by the director of operations, and that view is often different than the view wanted by a network engineer.

In spite of the importance of providing a holistic approach to APM, only about 15% of The Survey Respondents indicated that their organization's approach to APM was both top down and tightly coordinated.

***Only a small minority of IT organizations has a top down, tightly coordinated approach to APM.***

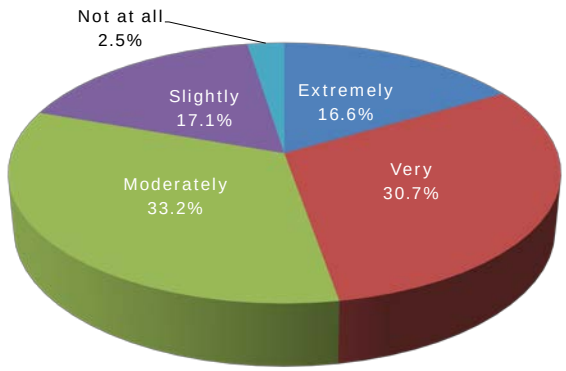
One of the reasons why it is important to get better at managing the user's experience was previously mentioned in the handbook. That reason being that in spite of all of the effort and resources that have gone into implementing IT management to date, it is the end user, and not the IT organization who typically is the first to notice when the performance of an application begins to degrade.

Monitoring actual user transactions in production environments provides valuable insight into the end-user experience and provides the basis for an IT organization to be able to quickly identify, prioritize, triage and resolve problems that can affect business processes.

To quantify the interest that IT organizations have in this task, The Survey Respondents were asked how important it was over the next year for their organization to get better at monitoring the end user's experience and behavior. Their responses are shown in **Figure 1**.

***Over the next year, getting better at monitoring the end user's experience and behavior is either very or extremely important to roughly half of all IT organizations.***

**Figure 1: Getting Better at Monitoring End User Behavior**




A holistic approach to APM must also address the following aspects of management:

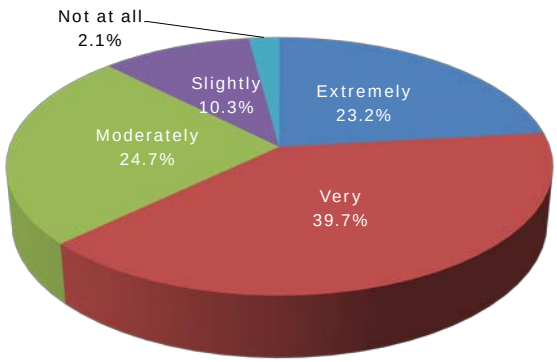
- The adoption of a system of service level agreements (SLAs) at levels that ensure effective business processes and user satisfaction for at least a handful of key applications.

- Automatic discovery of all the elements in the IT infrastructure that support each service. This functionality provides the basis for an IT organization to being able to create two-way mappings between the services and the supporting infrastructure components. These mappings, combined with event correlation and visualization, can facilitate root cause analysis, significantly reducing mean-time-to-repair.

The Survey Respondents were asked how important it was over the next year for their organization to get better at identifying the components of the IT infrastructure that support the company's critical business applications. Their responses are shown in **Figure 2**.

***Getting better at identifying the components of the IT infrastructure that support the company's critical business applications and services is one of the most important management tasks facing IT organizations.***

**Figure 2: Importance of Identifying Components of Critical Applications**



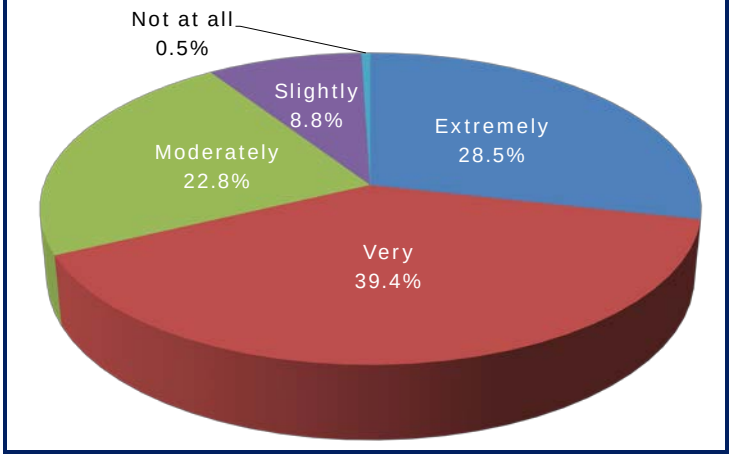
If IT organizations can effectively identify which components of the infrastructure support a particular application or service, monitoring can much more easily identify when services are about to begin to degrade due to problems in the infrastructure. As part of this monitoring, predictive techniques such as heuristic-based trending of software issues and infrastructure key performance indicators can be employed to identify and alert management of problems before they impact end users. In addition, outages and other incidents that generate alerts can be prioritized based on their potential business impact. Prioritization can be based on a number of factors including the affected business process and its value to the enterprise, the identity and number of users affected and the severity of the issue.

Once the components of the infrastructure that support a given application or service has been identified, triage and root cause analysis can be applied at both the application and the infrastructure levels. When applied directly to applications, triage and root cause analysis can identify application issues such as the depletion of threads and pooled resources, memory leaks or internal failures within a Java server or .NET server. At the infrastructure level, root cause analysis can determine the subsystem within the component that is causing the problem.

The Survey Respondents were asked how important it was over the next year for their organization to get better at rapidly identifying the causes of application degradation. Their responses are shown in **Figure 3**.

***Getting better at rapidly identifying the causes of application degradation is the most important management task facing IT organizations.***

**Figure 3: The Importance of Getting Better at Root Cause Analysis**




As part of an effective approach to APM, the automated generation of performance dashboards and historical reports allows both IT and business managers to gain insight into SLA compliance and performance trends. The insight that can be gleaned from these dashboards and reports can be used to enhance the way that IT supports key business processes, help the IT organization to perform better capacity and budget planning, and identify where the adoption of new technologies can further improve the optimization, control and management of application and service performance. Ideally, the dashboard is a single pane of glass that can be customized to suit different management roles; e.g., the individual contributors in the Network Operations Center, senior IT management as well as senior business management.

## Challenges for Application Management

Below is a discussion of some of the technical factors that complicate the ability of IT organizations to perform effective APM. While the technical factors present a significant challenge, an equally significant challenge is organizational – the difficulty of actually taking a top down, tightly integrated approach to APM.

## Server Virtualization

Server virtualization presents a number of challenges relative to APM. For example, the VMs that reside on a given physical server communicate with each other using a vSwitch within the server's hypervisor software. As discussed in the section of this handbook entitled Virtualization, unlike the typical physical switch, a vSwitch usually provides limited visibility for the traffic that is internal to the physical server. In addition, prior to virtualization, most server platforms were dedicated to a single application. With server virtualization, virtual machines share the server's CPU, memory and I/O resources. Over-subscription of VMs on a physical server can result in application performance problems due to factors such as limited CPU cycles or memory or I/O bottlenecks. One way to mitigate the impact of the over-subscription of VMs is to implement functionality such as VMotion[3] in an automated fashion. However, automated VMotion creates additional challenges.

While the problems discussed in the preceding paragraph can occur in a traditional physical server, they are more likely to occur in a virtualized server due to the consolidation of multiple applications onto a single shared physical server. In addition, as described in the section of this handbook entitled Virtualization, it is notably more difficult to troubleshoot a performance problem in a virtualized environment than it is in a traditional physical environment. That is why, as is also pointed out in that section of the handbook, half of the IT organizations consider it to be either very or extremely important over the next year for them to get better performing management tasks such as troubleshooting on a per-VM basis.

## Mobility

Another factor that is making APM more complex is that most IT organizations have to support a growing number of mobile employees. As described in the section of the handbook entitled Application and Service Delivery Challenges, at one time mobile workers tended to primarily access either recreational applications or business applications that were not very delay sensitive; e.g., email. However, mobile workers now need to access an increasingly wide range of business critical applications, many of which are delay sensitive. One of the issues associated with supporting mobile workers' access to delay sensitive, business critical applications is that because of the way that TCP functions, even the small amount of packet loss that is often associated with wireless networks results in a dramatic reduction in throughput. As such, there is a significant risk that an application that performs well when accessed over a wired network will run poorly when accessed over a wireless network.

The challenges associated with supporting mobility are why, as highlighted in the section of the handbook entitled Application and Service Delivery Challenges, two thirds of The Survey Respondents indicated that over the next year it is either moderately or very important for their IT organization to get better at managing the performance of applications delivered to mobile users.

## Cloud Computing

There are many ways that the adoption of cloud computing adds to the complexity of APM. For example, assume that the 4-tier application BizApp that was described in the section of this report that is entitled *Virtualization*, is moved to a cloud computing service provider's data center. Without the appropriate tools and processes it is impossible to tell in advance what

[3] VMWare.com VMotion

impact that move will have on application performance. However, the fact that BizApp will run on different servers, which are most likely virtualized, and is accessed over different WAN links than it had been previously, means that the application performance will be different. This lack of ability to understand in advance how a change in the IT environment will impact the performance of an application is one of the factors driving the need for Application Performance Engineering which is described below.

As was described in the section of the handbook entitled *Virtualization*, troubleshooting any performance degradation exhibited by BizApp is complex even if each tier of the application is hosted by an enterprise IT organization. However, if one or more tiers of the application are hosted by a CCSP troubleshooting becomes notably more complex because management data must now be gathered from multiple organizations.

## Port Hopping

As previously noted, identifying the applications and services that are running on a network is a critical part of managing application performance. TCP and UDP ports are frequently used by routers, firewalls and other network devices to identify the application that generated a particular packet. A well-known port serves as a contact point for a client to access a particular service over the network. For example, port 80 is the well-known port for HTTP data exchange and port 443 is the well-known port for secure HTTP exchanges via HTTPS.

Some applications have been designed to use port hopping to avoid detection and blocking by firewalls. Applications that do port hopping create significant management and security challenges. Two applications that often use port hopping are instant messaging (IM) and peer-to-peer (P2P) applications such as Skype.

## Instant Messaging

An example of a port-hopping instant messaging client is AOL's Instant Messenger (AIM). AOL has been assigned ports 5190 through 5193 for its Internet traffic, and AIM is typically configured to use these ports. As a result, network managers might well think that by blocking ports 5190 – 5193 they are blocking the use of AIM when in reality they are not. Analogously, network management might see that there is no traffic on ports 5190 – 5193 and assume that AIM is not being used. That may or may not be the case because if these ports are blocked AIM will use port 80 in an effort to circumvent the firewall via the Port 80 black hole described below.

## Peer-to-Peer Networks and Skype

A peer-to-peer computer network leverages the connectivity between the participants in a network. Unlike a typical client-server network where communication is typically to and from a central server along fixed connections, P2P nodes are generally connected via ad hoc connections. Such networks are useful for many purposes, including file sharing and IP telephony.

Skype is a peer-to-peer based IP telephony and IP video service developed by Skype Technologies SA – a company that Microsoft acquired. Many peer-to-peer applications, including Skype, change the port that they use each time they start. Consequently, there is no standard Skype port like there is a standard SIP port or a standard SMTP port. In addition, Skype is particularly adept at port-hopping with the aim of traversing enterprise firewalls. Once

inside the firewall, it then intentionally connects to other Skype clients. If one of those clients happens to be infected, then the machines that connect to it can be infected with no protection from the firewall. Moreover, because Skype has the ability to port-hop, it is much harder to detect anomalous behavior or configure network security devices to block the spread of the infection.

**The Port 80 Black Hole**

Many enterprise applications are accessed via browsers over port 80. Therefore, a firewall can't block port 80 without eliminating much of the traffic on which a business may depend. As mentioned, some applications will port-hop to port 80 when their normally assigned ports are blocked by a firewall. In addition, the port number 80 can't be used as a means of identifying individual web based enterprise applications and port 80 becomes a black hole unless firewalls and other devices are capable of deep packet inspection to identify Layer 7 application signatures.

***Lack of visibility into the traffic that transits port 80 is a significant management and security challenge for most IT organizations.***

The port 80 black hole can have four primary effects on an IT organization. It can cause increased:

- Difficulty in managing the performance of key business-critical, time-sensitive applications
- Vulnerability to security breaches
- Difficulty in complying with government and industry regulations
- Vulnerability to charges of copyright violation

## APM in the Private Enterprise Network[4]

Enterprise IT organizations can choose among several types of tools for monitoring and managing application performance over a private enterprise network. These include: application agents, monitoring of real and synthetic transactions, network flow and packet capture, analytics, and dashboard portals for the visualization of results.

At a high level, there are two basic classes of tools. The first class of tool monitors global parameters such as user response time or transaction completion time and provides alerts when thresholds are exceeded. These tools include agents on end user systems and monitoring appliances in the data center. The second class of tool supports triage by monitoring one or more of the components that make up the end-to-end path of the application. These tools include devices that capture application traffic at the flow and packet levels, agents on database, application, and web servers, as well as agents on various network elements.

[4] This refers to managing the performance of applications that are delivered over WAN services such as Frame Relay, ATM and MPLS.

The ultimate goal of APM is have a single screen that integrates the information from all of the tools in both categories. The idea being that a dashboard on the screen would indicate when user response time or transaction completion time begins to degrade. Then, within a few clicks, the administrator could determine which component of the infrastructure was causing the degradation and could also determine why that component of the infrastructure was causing degradation; e.g., high CPU utilization on a router.

Each type of individual tool has its strengths and weaknesses. For example, agents can supply the granular visibility that is required for complex troubleshooting but they represent an additional maintenance burden while also adding to the load on the servers and on the network. Monitoring appliances have more limited visibility, but they don't require modification of server configurations and don't add traffic to the network. Taking into consideration these trade-offs, IT organizations need to make tool decisions based on their goals for APM, their application and network environment as well as their existing infrastructure and network management vendors.

A complete discussion of APM tools and methodology is outside the scope of this section of the handbook. That said, the remainder of this section is devoted to the following topics that are of particular importance for APM within the private enterprise network:

- **Application Performance Engineering** that deals with the processes of optimizing the performance of applications over their lifecycles.
- **End-to-End Visibility** of all aspects of all the infrastructure components that can have an effect of application performance.
- **Route Analytics** that deals with mitigating the logical issues within the routed IP network that can negatively impact application performance.

## Application Performance Engineering

Ideally the issue of application performance would be addressed at all stages of an application's lifecycle, including multiple iterations through the design/implement/test/operate phases as the application versions are evolved to meet changing requirements. However, as discussed in a preceding section of the handbook, the vast majority of IT organizations don't have any insight into the performance of an application until after the application is fully developed and deployed. In addition, the vast majority of IT organizations have little to no insight into how a change in the infrastructure, such as implementing server virtualization, will impact application performance prior to implementing the change.

***Application Performance Engineering (APE) is the practice of first designing for acceptable application performance and then testing, measuring and tuning performance throughout the application lifecycle.***

During the operational, or production phase of the lifecycle, APM is used to monitor, diagnose, and report on application performance. APM and APE are therefore highly complementary disciplines. For example, once an APM solution has identified that an application in production is experiencing systemic performance problems, an APE solution can be used to identify the root cause of the problem and to evaluate alternative solutions. Possible solutions include modifying the application code or improving application performance by making changes in the supporting infrastructure, such as implementing more highly performing servers or deploying

WAN Optimization Controllers (WOCs). Throughout this section of the handbook, implementing products such as WOCs will be referred to as a Network and Application Optimization (NAO) solution. Independent of which remedial option the IT organization takes, the goal of APE can be realized – performance bottlenecks are identified, root causes are determined, alternative remedies are analyzed and bottlenecks are eliminated.

An IT organization could decide to ignore APE and just implement NAO in a reactive fashion in an attempt to eliminate the sources of the degraded application performance. Since this approach is based on the faulty assumption that NAO will resolve all performance problems, this approach is risky. This approach also tends to alienate the company's business unit managers whose business processes are negatively impacted by the degraded application performance that is not resolved until either WOCs are successfully deployed or some other solution is found. A more effective approach was described in the preceding paragraph. This approach calls for NAO to be a key component of APE – giving IT organizations another option to proactively eliminate performance problems before they impact key business processes.

The key components of APE are described below. The components are not typically performed in a sequential fashion, but in an iterative fashion. For example, as a result of performing testing and analysis, an IT organization may negotiate with the company's business unit managers to relax the previously established performance objectives.

- **Setting Performance Objectives**
  This involves establishing metrics for objectives such as user response time, transaction completion time and throughput. A complex application or service, such as unified communications, is comprised of several modules and typically different objectives need to be established for each module.

- **Discovery**
  Performance modeling and testing should be based on discovering and gaining a full understanding of the topology and other characteristics of the production network.

- **Performance Modeling**
  APE modeling focuses on creating the specific usage scenarios to be tested as well as on identifying the performance objectives for each scenario. A secondary focus is to identify the maximum utilization of IT resources (e.g., CPU, memory, disk I/O) and the metrics that need to be collected when running the tests.

- **Performance Testing and Analysis**
  Test tools can be configured to mimic the production network and supporting infrastructure, as well as to simulate user demand. Using this test environment, the current design of the application can be tested in each of the usage scenarios against the various performance objectives. The ultimate test, however, is measured performance in the actual production network or in a test environment that very closely mimics the actual production environment.

- **Optimization**
  Optimization is achieved by identifying design alternatives that could improve the performance of the application and by redoing the performance testing and analysis to quantify the impact of the design alternatives. In conjunction with the testing, an ROI analysis can be performed to facilitate cross-discipline discussion of the tradeoffs

between business objectives, performance objectives, and cost. This component of APE is one of the key ways that APE enables an IT organization to build better relationships with the company's business unit managers.

## End-to-End Visibility

The IT industry uses the phrase end-to-end visibility in various ways. Given that one of this handbook's major themes is that IT organizations need to implement an application-delivery function that focuses directly on applications and not on the individual components of the IT infrastructure, this handbook will use the following definition of end-to-end visibility:

***End-to-end visibility refers to the ability of the IT organization to examine every component of IT that impacts communications once users hit ENTER or click the mouse button until they receive a response back from the application.***

End-to-end visibility is one of the cornerstones of assuring acceptable application performance. This functionality is important because it:

- Provides the information that allows IT organizations to notice application performance degradation before the end user does.

- Identifies the symptoms of the degradation and as a result enables the IT organization to reduce the amount of time it takes to identify and remove the causes of the degraded application performance.

- Facilitates making intelligent decisions and getting buy-in from other impacted groups. For example, end-to-end visibility provides the hard data that enables an IT organization to know that it needs to add bandwidth or redesign some of the components of the infrastructure because the volume of traffic associated with the company's sales order tracking application has increased dramatically. It also positions the IT organization to manage the recreational use of the network.

- Allows the IT organization to measure the performance of a critical application before, during and after a change is made. These changes could be infrastructure upgrades, configuration changes or the adoption of a cloud computing delivery model. As a result, the IT organization is in a position both to determine if the change has had a negative impact and to isolate the source of the problem so it can fix the problem quickly.

Visibility can enable better cross-functional collaboration if two criteria are met. One criterion is that all members of the IT organization use the same tool or set of tools. The second criterion is that the tool(s) are detailed and accurate enough to identify the sources of application degradation. One factor that complicates achieving this goal is that so many tools from so many types of vendors (e.g., APM, NAO) all claim to provide the necessary visibility.

Providing detailed end-to-end visibility is difficult due to the complexity and heterogeneity of the typical enterprise network. The typical enterprise network, for example, is comprised of switches and routers, access points, firewalls, ADCs, WOCs, intrusion detection and intrusion prevention appliances from a wide range of vendors. An end-to-end monitoring solution must profile traffic in a manner that reflects not only the physical network but also the logical flows of

applications, and must be able to do this regardless of the vendors who supply the components or the physical topology of the network.

The sub-section of the handbook entitled *Virtualization* highlighted a visibility challenge created by server virtualization. That problem is that in most cases once a server is virtualized the IT organization loses visibility into the inter-VM traffic on a given server. There are a number of solutions for this problem. One of these solutions is based on configuring one of the ports on the virtual switch inside the server as a SPAN port or mirror port. This allows a monitor to capture flow and packet information within the physical server. The monitoring device can be a virtual appliance installed on the physical server. Transaction and response time monitors are also available as virtual appliances. While changes in the virtual topology can be gleaned from flow analysis, a more direct approach is for the APM tool to access data in the hypervisor's management system via supported APIs. Gathering data from this source also provides access to granular performance information such as a VM's utilization of allocated CPU and memory resources.

When implementing techniques to gain end-to-end visibility, IT organizations have easy access to management data from both SNMP MIBs and from NetFlow. IT organizations also have the option of deploying either dedicated instrumentation or software agents to gain a more detailed view into the types of applications listed below. An end-to-end visibility solution should be able to identify:

- Well-known application layer protocols; e.g. FTP, Telnet, HTTPS and SSH.
- Services, where a service is comprised of multiple inter-related applications.
- Applications provided by a third party; e.g., Oracle, Microsoft, SAP.
- Applications that are not based on IP; e.g., applications based on IPX or DECnet.
- Custom or homegrown applications.
- Web-based applications.
- Multimedia applications.

Relative to choosing an end-to-end visibility solution, other selection criteria include the ability to:

- Scale as the size of the network and the number of applications grows.
- Add minimum management traffic overhead.
- Support granular data collection.
- Capture performance data as well as events such as a fault.
- Support a wide range of topologies both in the access, distribution and core components of the network as well as in the storage area networks.

- Support real-time and historical analysis.
- Integrate with other management systems.
- Support flexible aggregation of collected information.
- Provide visibility into complex network configurations such as load-balanced or fault-tolerant, multi-channel links.
- Support the monitoring of real-time traffic.
- Generate and monitor synthetic transactions.

## Route Analytics

### Background

The use of route analytics for planning purposes was discussed in the preceding section of the handbook. This section of the handbook will expand on the use of route analytics for operations.

One of the many strengths of the Internet Protocol (IP) is its distributed intelligence. For example, routers exchange reachability information with each other via a routing protocol such as OSPF (Open Shortest Path First). Based on this information, each router makes its own decision about how to forward a packet. This distributed intelligence is both a strength and a weakness of IP. In particular, while each router makes its own forwarding decision, there is no single repository of routing information in the network.

The lack of a single repository of routing information is an issue because routing tables are automatically updated and the path that traffic takes to go from point A to point B may change on a regular basis. These changes may be precipitated by a manual process such as adding a router to the network, the mis-configuration of a router or by an automated process such as automatically routing around a failure. In this latter case, the rate of change might be particularly difficult to diagnose if there is an intermittent problem causing a flurry of routing changes typically referred to as route flapping. Among the many problems created by route flapping is that it consumes a lot of the processing power of the routers and hence degrades their performance.

The variability of how the network delivers application traffic across its multiple paths in a traditional IT environment can undermine the fundamental assumptions that organizations count on to support many other aspects of application delivery. For example, routing instabilities can cause packet loss, latency and jitter on otherwise properly configured networks. In addition, alternative paths might not be properly configured for QoS. As a result, applications perform poorly after a failure. Most importantly, configuration errors that occur during routine network changes can cause a wide range of problems that impact application delivery. These configuration errors can be detected if planned network changes can be simulated against the production network.

As previously noted in this handbook, the majority of IT organizations have already implemented server virtualization and the amount of server virtualization is expected to increase over the next

year. Once an IT organization has implemented server virtualization, or a private cloud computing solution that includes server virtualization, VMs can be transferred without service interruption from a given physical server to a different physical server. This can make it difficult for the network operations team to know the location of an application at any given point in time – a fact that makes troubleshooting a problem that much more difficult.

To exemplify a related management challenge, assume that an IT organization has implemented a type of hybrid cloud computing solution whereby the IT organization hosts the application and data base tiers in one of their data centers and that the relevant servers have been virtualized. Further assume that a CCSP hosts the application's web tier and that all of the CCSP's physical servers have been virtualized. All of the users access the application over the Internet and the connectivity between the web server layer and the application server layer is provided by an MPLS service.

Since the web, application and database tiers can be moved, either dynamically or manually, it is extremely difficult at any point in time for the IT operations organization to know the exact routing between the user and the web tier, between the Web tier and the application tier or between the application tier and the database tier. This difficulty is compounded by that fact that as previously discussed, not only does the location of the tiers of the application change, but the path that traffic takes to go from point A to point B also changes regularly.

The dynamic movement of VMs will increase over the next few years in part because organizations will increase their use of virtualization and cloud computing and in part because organizations will begin to deploy techniques such as cloud bursting. Cloud bursting refers to taking an application that currently runs in a data center controlled by an IT organization and dynamically deploying that application and the subtending storage in a data center controlled by a CCSP. Techniques such as cloud bursting will enable organizations to support peak demands while only deploying enough IT infrastructure internally to support the average demand. These techniques, however, will further complicate the task of understanding how traffic is routed end-to-end through a complex, meshed network.

***The operational challenges that are created due to a lack of insight into the router layer are greatly exacerbated by the adoption of server virtualization and cloud computing.***

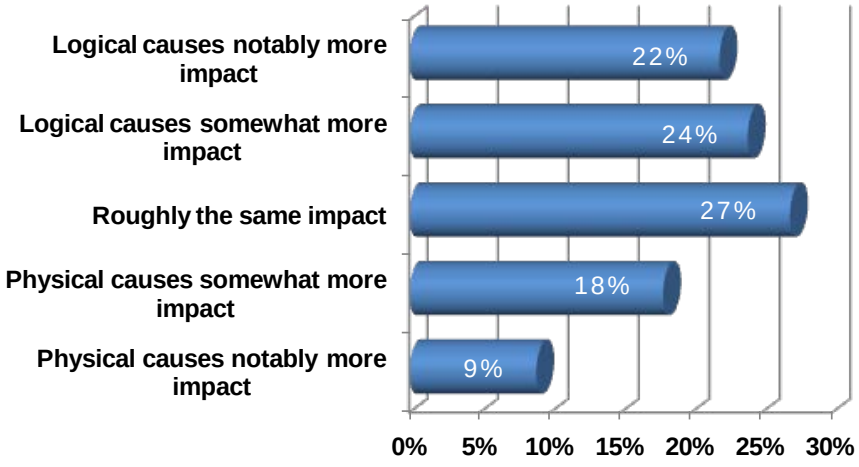
## Logical vs. Physical Factors

Factors such as route flapping can be classified as logical as compared to a device specific factor such as a link outage, which is considered to be a physical factor. Both logical and physical factors impact application performance. In simple networks, such as small hub and spoke networks, logical factors are typically not a significant source of application degradation. However, in large complex networks that is not the case.

To quantify the relative impact of logical and physical factors, The Survey Respondents were asked two questions.

One question asked The Survey Respondents to indicate the relative impact of logical and physical factors on the business disruption they cause. Their answers are shown in **Figure 4**.

**Figure 4: Impact of Logical and Physical Factors on the Business**

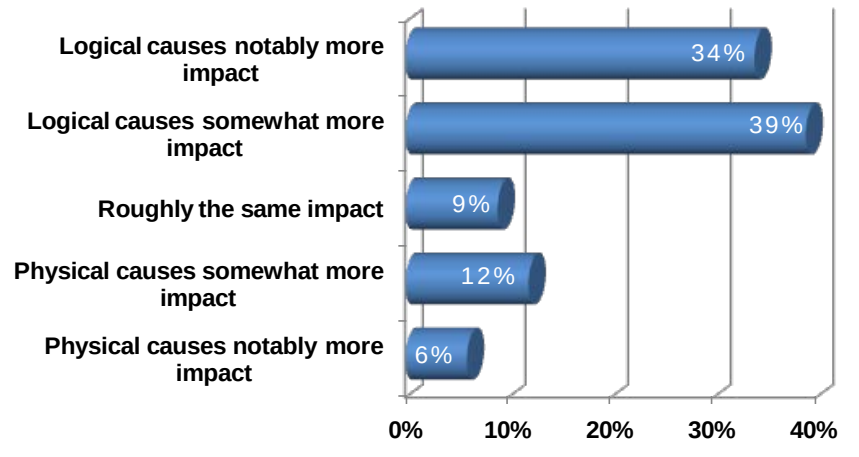



***In the vast majority of cases, logical factors cause as much or more business disruption than do physical factors.***

The other question asked The Survey Respondents to indicate the relative amount of time it takes to troubleshoot and repair a physical error vs. a logical error. Their answers are shown in **Figure 5**.

***In the vast majority of instances, logical errors take either somewhat more or notably more time to troubleshoot and repair than do physical errors.***

**Figure 5: Impact of Logical and Physical Factors on Troubleshooting**




SNMP-based management systems can discover and display the individual network elements and their physical or Layer 2 topology. However, these systems cannot identify the actual routes packets take as they transit the network. As such, SNMP-based systems cannot easily identify problems such as route flaps or mis-configurations.

As noted in the preceding section, the goal of route analytics is to provide visibility, analysis and diagnosis of the issues that occur at the routing layer. A route analytics solution achieves this goal by providing an understanding of precisely how IP networks deliver application traffic. This

requires the creation and maintenance of a map of network-wide routes and of all of the IP traffic flows that traverse these routes. This in turn means that a route analytics solution must be able to record every change in the traffic paths as controlled and notified by IP routing protocols.

By integrating the information about the network routes and the traffic that flows over those routes, a route analytics solution can provide information about the volume, application composition and class of service (CoS) of traffic on all routes and all individual links. This network-wide, routing and traffic intelligence serves as the basis for:

- Real-time monitoring of the network's Layer 3 operations from the network's point of view.
- Historical analysis of routing and traffic behavior as well as for performing a root causes analysis.
- Modeling of routing and traffic changes and simulating post-change behavior.

Criteria to evaluate a route analytics solution is the ability of the solution to:

- Listen to and participate in the routing protocol exchanges between routers as they communicate with each other.
- Compute a real-time, network-wide routing map. This is similar in concept to the task performed by individual routers to create their forwarding tables. However, in this case it is computed for all routers.
- Map Netflow traffic data, including application composition, across all paths and links in the map.
- Monitor and display routing topology and traffic flow changes as they happen.
- Detect and alert on routing events or failures as routers announce them, and report on correlated traffic impact.
- Correlate routing events with other information, such as performance data, to identify the underlying cause and effect.
- Record, analyze and report on historical routing and traffic events and trends.
- Simulate the impact of routing or traffic changes on the production network.

Another criterion that an IT organization should look at when selecting a route analytics solution is the breadth of routing protocol coverage. For example, based on the environment, the IT organization might need the solution to support protocols such as OSPF, IS-IS, EIGRP, BGP and MPLS VPNs. One more criterion is that the solution should be able to collect data and correlate integrated routing and Netflow traffic flow data. Ideally, this data is collected and reported on in a continuous real-time fashion and is also stored in such a way that it is possible to generate meaningful reports that provide an historical perspective on the performance of the network. The solution should also be aware of both application and CoS issues and be able to

integrate with other network management components. In particular, a route analytics solution should be capable of being integrated with network-agnostic application performance management tools that look at the endpoint computers that are clients of the network, as well as with traditional network management solutions that provide insight into specific points in the network; i.e., devices, interfaces, and links.

## APM in Public and Hybrid Clouds

As is widely known, IT organization have begun to make significant use of public and hybrid cloud computing solutions and the use of those solutions is expected to increase significantly. Once enterprise applications are partially or completely hosted outside of private data centers, IT organizations will need to make some adjustments in their approach to APM. In particular, public clouds have a significant impact on each the topics discussed in the preceding section.

- ***APE***
  While an enterprise IT organization might hope that a SaaS provider would use APE as part of developing their application, they typically can't cause that to happen. IT organizations can, however, use APE to quantify the impact of taking an application, or piece of an application, that is currently housed internally and hosting it externally. IT organizations can also use APE for other cloud related activities, such as quantifying the impact on the performance of a SaaS based application if a change is made within the enterprise. For example, APE can be used to measure the impact of providing mobile users with access to a SaaS-based application that is currently being used by employees in branch offices.

- ***End-to-End Visibility***
  The visibility necessary for effective APM can be compromised by the dynamic nature of cloud environments and by the difficulty of extending the enterprise monitoring solutions for application servers, Web servers, databases into a public IaaS cloud data center. Part of this challenge is that many IaaS providers have an infrastructure that has often been optimized based on simplicity, homogeneity and proprietary extensions to open source software.

- ***Route Analytics***
  As noted in the preceding section, both hosting enterprise assets at a CCSP's premise, and using services provided by a CCSP creates a more complex network topology. This fact combined with the potential for the dynamic movement of those assets and services increases the probability of a logical error. As such, the adoption of cloud based service increases the need for route analytics.

There are a number of possible ways that an IT organization can adjust their APM strategies in order to accommodate accessing services hosted by a CCSP. These include:

- Extend the Enterprise APM Monitoring solutions into the public cloud using agents on virtual servers and by using virtual appliances. This option assumes that the CCSP offers the ability to install multiple virtual appliances (e.g., APM monitors, WOCs, and ADCs) and to configure the virtual switches to accommodate these devices.

- Focus on CCSPs that offer either cloud resource monitoring or APM as a service as described in the section of the handbook entitled Cloud Networking Services. Basic

cloud monitoring can provide visibility into resource utilization, operational performance, and overall demand patterns. This includes providing metrics such as CPU utilization, disk reads and writes and network traffic. The value of cloud monitoring is increased where it is tied to other capabilities such as automated provisioning of instances to maintain high availability and the elastic scaling of capacity to satisfy demand spikes. A possible issue with this option is integrating the cloud monitoring and enterprise monitoring and APM solutions.

- Increase the focus on service delivery and transaction performance by supplementing existing APM solutions with capabilities that provide an outside-in service delivery view from the perspective of a client accessing enterprise applications or cloud applications over the Internet or mobile networks. Synthetic transactions against application resources located in public clouds are very useful when other forms of instrumentation cannot be deployed. One option for synthetic transaction monitoring of web applications is a third party performance monitoring service with end user agents distributed among numerous global ISPs and mobile networks.

## Security

The section of this handbook that is entitled “First Generation Application and Service Delivery Challenges” referenced a number of industry reports in order to describe the current security environment and to identify the types of attacks that are becoming increasingly common.

For example, that section of the handbook referenced IBM’s X-Force 2011 Trend and Risk Report[5]. Some of the key observations made in that report are:

- **Mobile Devices**
  The report stated that in 2011 there was a 19 percent increase over 2010 in the number of exploits publicly released that can be used to target mobile devices such as those that are associated with the BYOD movement. The report added that there are many mobile devices in consumers' hands that have unpatched vulnerabilities to publicly released exploits, creating an opportunity for attackers.

- **Social Media**
  With the widespread adoption of social media platforms and social technologies, this area has become a target of attacker activity. The IBM report commented on a surge in phishing emails impersonating social media sites and added that the amount of information people are offering in social networks about their personal and professional lives has begun to play a role in pre-attack intelligence gathering for the infiltration of public and private sector computing networks.

- **Cloud Computing**
  The report stated that there were many high profile cloud breaches affecting well-known organizations and large populations of their customers. IBM recommended that IT security staff should carefully consider which workloads are sent to third-party cloud providers and what should be kept in-house due to the sensitivity of data. The IBM X-Force report also noted that the most effective means for managing security in the cloud may be through Service Level Agreements (SLAs) and that IT organizations should pay careful consideration should be given to ownership, access management, governance and termination when crafting SLAs.

That section also referenced Blue Coat Systems’ 2012 Web Security Report[6]. According to the Blue Coat report, “In 2011, malnets emerged as the next evolution in the threat landscape. These infrastructures last beyond any one attack, allowing cybercriminals to quickly adapt to new vulnerabilities and repeatedly launch malware attacks. By exploiting popular places on the Internet, such as search engines, social networking and email, malnets have become very adept at infecting many users with little added investment.” That report also noted the increasing importance of social networking and stated that, “Since 2009, social networking has increasingly eclipsed web-based email as a method of communications.” and that, “Now, social networking is moving into a new phase in which an individual site is a self-contained web environment for many users – effectively an Internet within an Internet.”

[5] X-Force 2011 Trend and Risk Report
[6] http://www.bluecoat.com/sites/default/files/documents/files/BC_2012_Security_Report-v1i-optimized.pdf

This section of the handbook will discuss the technologies and services that IT organizations can use to respond to these security challenges.

## How IT Organizations are Implementing Security

As previously described in this handbook, the security landscape has changed dramatically in the last few years. In the very recent past, the typical security hacker worked alone, relied on un-sophisticated techniques such as dumpster diving, and was typically motivated by the desire to read about their hack in the trade press. In the current environment, sophisticated cyber criminals have access to malware networks and R&D labs and can use these resources to launch attacks whose goal is usually to make money for the attacker. National governments and politically active hackers (hacktivists) are engaging in cyber warfare for a variety of politically motivated reasons. Unfortunately terms such as Distributed Denial-of-Service (DDoS), Advanced Persistent Attacks (APTs) and SQL injections attacks are becoming common used as illicit activity continues.

In many aspects security is both a first and a second-generation application and service delivery challenge and it will remain a significant challenge for the foreseeable future. Rapid changes in IT, such as those created by the adoption of cloud computing, social networking and the new generation of mobile devices, combined with the ongoing evolution of regulations pose a spate of new challenges for IT security systems and policies in much the same manner that they present challenges to the IT infrastructure.

IT security systems and policies have evolved and developed around the traditional application delivery architecture in which branch offices are connected to application servers in a central corporate data centers. In this architecture, the central corporate data center is a natural location to implement IT security systems and policies that provide layered defenses as well a single, cost efficient location for a variety of IT security functions. With the adoption of public cloud computing, applications and services are moving out of the central corporate data center and there is no longer a convenient single location for security policies and systems.
IT security systems and policies have traditionally distinguished between people who were using IT services for work versus those who were using it for personal use. The use of an employer provided laptop was subject to the employer's IT security policies and systems. In this environment, the use that employees made of personal laptops was generally outside of the corporate IT security policy. With the arrival of smartphones and tablet computers, the ownership, operating systems and security capabilities of the end user devices have changed radically. IT security policies and standards that were developed for PCs are no longer effective nor optimal with these devices. Most corporations have embraced the BYOD movement and end users are less willing to accept strict corporate security policies on devices they own. Additionally, strict separation of work and personal usage for security on an employee owned device is impractical.

The demands of governments, industry and customers have historically shaped IT security systems and policies. The wide diversity of organizations that create regulations and standards can lead to conflicts. For example, law enforcement requires access to network communications (Communications Assistance for Law Enforcement Act – CALEA) which may in turn force the creation of locations in the network that do not comply with the encryption requirements of other standards (e.g. Health Insurance Portability Accountability Act – HIPPA). In order to determine how IT organizations are responding to the traditional and emerging security challenges, The Survey Respondents were asked a series of questions. For example,

to get a high level view of how IT organizations are providing security, The Survey Respondents were asked to indicate which of a number of network security systems their organization supports. The Survey Respondents were asked to check all of the alternatives that applied in their environment. Their responses are shown in **Table 1**.

**Table 1: The Network Security Systems in Use**

| Network Security Systems | Percentage |
|---|---|
| Remote Access VPN | 86.30% |
| Network Access Control | 73.50% |
| Intrusion Detection/Protection Systems (IDS/IPS) | 65.70% |
| Next Generation Firewalls (Firewall+IPS+Application Control) | 56.90% |
| Secure Web Gateways | 46.10% |
| Web Application and/or XML Firewalls | 36.30% |
| Mobile Device Security/Protection | 36.30% |
| Security Information Event Management | 31.40% |
| Data Loss Prevention | 24.50% |
| Password Vault Systems (either local or portal based) | 12.70% |
| SAML or WS-Federation Federated Access Control | 8.80% |

One obvious conclusion that can be drawn from **Table 1** is that IT organizations use a wide variety of network security systems. A slightly less obvious conclusion is that on average, IT organizations use 4.8 of the network security systems listed in the preceding table.

The Survey Respondents were asked to indicate the approach that best describes how their company uses data classification to create a comprehensive IT security environment. Their responses are shown in **Table 2**.

**Table 2: Approach to Comprehensive IT Security**

| Approach | Percentage |
|---|---|
| We have a data classification policy and it is used to determine application access/authentication, network and end user device security requirements. | 42.90% |
| We do not have a data classification policy. | 33.00% |
| We have a data classification policy and it is used to determine application security requirements. | 13.20% |
| We have a data classification policy, but it is not used nor enforced. | 11.00% |

The data in **Table 2** represents a classic good news/bad news situation. The good news is that the majority of IT organizations have a data classification policy that they use to determine requirements. The bad news is that 44% of IT organizations either don't have a data classification policy or they have one that isn't used or enforced.

In order to understand how IT organizations are responding to the BYOD movement, The Survey Respondents were asked, "If your organization does allow employee owned devices to connect to your network, please indicate which of the following alternatives are used to register employee owned devices and load authentication (e.g. certificate/private key) data onto those devices before they are allowed to connect to your company's network." The Survey Respondents were asked to check all of the alternatives that applied in their environment. Their responses are shown in **Table 3**.

**Table 3: Alternatives to Support Employee Owned Devices**

| Alternative | Percentage |
|---|---|
| Employees must install a VPN client on their devices for network access | 53.90% |
| IT Administrator and/or Service Desk must register employee owned device for network access | 47.40% |
| Employees can self-register their devices for network access | 28.90% |
| Employees must generate and/or load X.509 certificates & private keys network access | 13.20% |
| Employees must install a token authentication app on their devices for network access | 10.50% |

The data in **Table 3** indicates that while using a VPN is the most common technique that a wide range of techniques are used. VPN's popularity comes in part from the fact that remote access VPN solutions implemented on new generation mobile devices have various capabilities to enforce security policies when connecting to the corporate network. Popular security checks include ensuring that a screen password is present, that anti-virus software is present and is up to date, that there is not rogue software on the device and that the operating system has not been modified.

Two different approaches have emerged to protect against lost devices. For the traditional PC, full disk encryption is typically used to protect data if the PC is lost or stolen. However, on new generation mobile devices, remote erase solutions are typically used to protect data. New generation mobile devices with smaller displays are often used more for content reading rather than content creation. As screen sizes and resolution improves, this situation may change. In order to understand how IT organizations have implemented full disk encryption, The Survey Respondents were asked to indicate which alternatives their organization implements relative to using full disk encryption on laptops and desktop PCs. Their responses are shown in **Table 4**.

**Table 4: Techniques for Implementing Full Disk Encryption**

| Alternative | Percentage |
|---|---|
| We do not use full disk encryption on PCs. | 52.5% |
| We use software based disk encryption on PCs. | 49.5% |
| We use hardware based self-encrypting rotating drives on PCs. | 6.1% |
| We use hardware based self-encrypting Solid State Drives on PCs. | 6.1% |

The data in **Table 4** indicates that just over half of all IT organizations don't use full disk encryption on PCs. The data also indicates that those IT organizations that do use full disk

encryption do so by using a software solution and that a small percentage of IT organizations use multiple techniques.

The Survey Respondents were asked to indicate the approach that best describes their company's approach to Identity and Access Management (IAM). Their responses are shown in **Table 5**.

**Table 5: How IAM is Implemented**

| Approach | Percentage |
|---|---|
| We do not have a formal IAM program. | 36.6% |
| We have an IAM program, but it only partially manages identities, entitlements and policies/rules for internal users. | 25.8% |
| We have an IAM program and it manages identities, entitlements and policies/rules for all internal users. | 20.4% |
| We have an IAM program and it manages identities, entitlements and policies/rules for end users for internal, supplier, business partner and customers. | 17.2% |

The data in **Table 5** indicates that only a minority of IT organizations has a IAM program that has broad applicability.

The Survey Respondents were asked to indicate how their company approaches the governance of network and application security. Their responses are shown in **Table 6**.

**Table 6: Governance Models in Use**

| Approach | Percentage |
|---|---|
| Network Security and Application Security are funded, architected, designed and operated together. | 46.9% |
| Network Security and Application Security are funded, architected, designed and operated separately. | 30.2% |
| Network Security and Application Security are funded jointly, but architected, designed and operated separately. | 22.9% |

The data in **Table 6** indicates that in the majority of instances, network security and application security are architected, designed and operated separately.

## Cloud-Based Security

The section of this handbook that focused on the emerging application and service delivery challenges presented the results of a survey in which The Survey Respondents were asked how likely it was over the next year that their company would acquire a traditional IT service from an IaaS provider. Their responses are shown in **Table 7**.

**Table 7: Interest in Obtaining IT Services as a Cloud-Based Service**

| | Will Not Happen | Might Happen | 50/50 Chance | Will Likely Happen | Will Happen |
|---|---|---|---|---|---|
| VoIP | 32.6% | 18.6% | 15.3% | 13.5% | 20.0% |
| Unified Communications | 30.2% | 22.8% | 20.5% | 14.9% | 11.6% |
| Security | 42.6% | 17.1% | 14.4% | 11.6% | 14.4% |
| Network and Application Optimization | 32.1% | 28.8% | 16.0% | 14.6% | 8.5% |
| Network Management | 41.4% | 22.3% | 13.5% | 13.5% | 9.3% |
| Application Performance Management | 37.9% | 26.5% | 15.6% | 11.4% | 8.5% |
| Virtual Desktops | 38.8% | 28.0% | 15.9% | 12.1% | 5.1% |

As shown in **Table 7**, the interest shown by The Survey Respondents in obtaining security as a Cloud-based service is bimodal. When looking just at the percentage of The Survey Respondents that indicated that it either will happen or will likely happen, security is one of the most likely services that IT organizations will acquire from a CCSP. However, a higher percentage (42.6%) of The Survey Respondents indicated that they will not acquire security from a CCSP than made that indication for any other form of IT service listed in the survey.

One way that a Cloud-based Security Service (CBSS) could provide value is if it provides protection against the growing number of malware attacks. To effectively protect against malware attacks, a CBSS should be able to identify suspicious content or sites that are either suspicious or are known to distribute malware. In order to be effective, a CBSS that provides Web content filtering or malware protection needs a source of intellectual capital that identifies known and suspected vulnerabilities. This source needs to be both dynamic and as extensive as possible.

One part of the value proposition of a CBSS that provides security functionality is the previously discussed value proposition of any cloud based service. For example, a security focused CBSS reduces the investment in security that an organization would have to make. In addition, a security focused CBSS reduces the amount of time it takes to deploy new functionality. The speed at which changes can be made to a CBSS adds value in a variety of situations, including providing better protection against zero-day attacks[7]. Another part of the value proposition of a security focused CBSS is that unlike a traditional security solution that relies on the

[7] http://en.wikipedia.org/wiki/Zero-day_attack

implementation of a hardware based proxy, a CBSS can also protect mobile workers. The CBSS does this by leveraging functionality that it provides at its POPs as well as functionality in a software agent that is deployed on each mobile device. The use of a Cloud-based solution to provide mobile device management and security was discussed previously in this section.

In many instances, the best security solution is a hybrid solution that combines traditional on-premise functionality with one or more Cloud-based solutions. For example, in many cases IT organizations already have functionality such as web filtering or malware protection deployed in CPE at some of their sites. In this case, the IT organization may choose to implement a CBSS just to protect the sites that don't have security functionality already implemented and/or to protect the organization's mobile workers. Alternatively, an organization may choose to implement security functionality in CPE at all of their sites and to also utilize a CBSS as part of a defense in depth strategy.

Other situations in which a CBSS can serve to either be the only source of security functionality, or to compliment CPE based implementations include cloud-based firewall and cloud-based IPS services. Such a service should support equipment from the leading vendors. Given the previously mentioned importance of hybrid solutions, the service should allow for flexibility in terms of whether the security functionality is provided in the cloud or from CPE as well as for flexibility in terms of who manages the functionality – a CCSP or the enterprise IT organization.

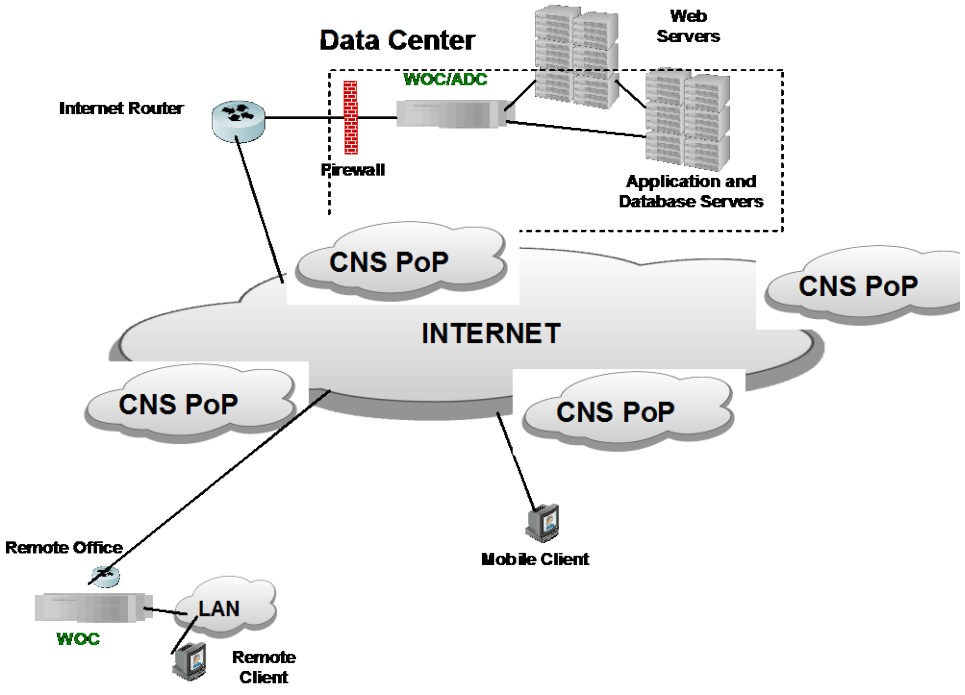
In addition to the specific security functionality provided by the CBSS, the CBSS should also:

- Provide predictive analytics whereby the CBSS can diagnose the vast majority of potential enterprise network and security issues before they can impact network health.
- Incorporate expertise, tools, and processes to ensure that the service that is provided can meet auditing standards such as SAS-70 as well as industry standards such as ITIL.
- Integrate audit and compliance tools that provide the necessary event-correlation capabilities and reporting to ensure that the service meets compliance requirements such as Sarbanes-Oxley, HIPAA, GLB and PCI.
- Provide the real-time notification of security events.

## Web Application Firewall Services

The section of this report entitled *Network and Application Optimization*, discussed how a Cloud-based service, such as the one shown in **Figure 6**, can be used to optimize the performance of the Internet. As will be discussed in this sub-section of the handbook, that same type of service can also provide security functionality.

**Figure 6: Internet Based Security Functionality**




## Role of a Traditional Firewall: Protect the Perimeter

Roughly twenty years ago IT organizations began to implement the first generation of network firewalls, which were referred to as packet filters. These devices were placed at the perimeter of the organization with the hope that they would prevent malicious activities from causing harm to the organization.

Today most network firewalls are based on stateful inspection. A stateful firewall holds in memory attributes of each connection. These attributes include such details as the IP addresses and ports involved in the connection and the sequence numbers of the packets traversing the connection. One of the weaknesses associated with network firewalls is that they are typically configured to open up ports 80 and 443 in order to allow passage of all HTTP and SSL traffic. Given that ports 80 and 443 are generally configured to be open, this form of perimeter defense is porous at best.

Whereas network firewalls are focused on parameters such as IP address and port numbers, a more recent class of firewall, referred to as a Web application firewall, analyzes messages at layer 7 of the OSI model. Web application firewalls are typically deployed as a hardware appliance and they sit behind the network firewall and in front of the Web servers. They look for violations in the organization's established security policy. For example, the firewall may look for abnormal behavior, or signs of a known attack. It may also be configured to block specified content, such as certain websites or attempts to exploit known security vulnerabilities. Because of their ability to perform deep packet inspection at layer 7 of the OSI model, a Web application firewall provides a level of security that cannot be provided by a network firewall.

## Defense in Depth: The Role of a Web Application Firewall Service

There are fundamental flaws with an approach to security that focuses only on the perimeter of the organization. To overcome these flaws, most IT organizations have moved to an approach to security that is typically referred to as *defense in depth*. The concept of defense in depth is not new. What is new in the current environment is the use of a CBSS to provide Web application firewall functionality that is distributed throughout the Internet. This means that Web application functionality is close to the source of security attacks and hence can prevent many security attacks from reaching the organization.

In the current environment, high-end DDoS attacks can generate 100 Gbps of traffic or more[8]. Attacks of this magnitude cannot be prevented by onsite solutions. They can, however, be prevented by utilizing a CBSS that includes security functionality analogous to what is provided by a Web application firewall and that can identify and mitigate the DDoS-related traffic close to attack traffic origin.

There is a wide range of ways that a DDoS attack can cause harm to an organization in a number of ways, including the:

- Consumption of computational resources, such as bandwidth, disk space, or processor time.
- Disruption of configuration information, such as routing information.
- Disruption of state information, such as the unsolicited resetting of TCP sessions.
- Disruption of physical network components.
- Obstructing the communication media between the intended users and the victim so that they can no longer communicate adequately.

Because there are a variety of possible DDoS attacks, IT organizations need to implement a variety of defense in depth techniques. This includes:

- **Minimizing the points of vulnerability**
  If an organization has most or all of its important assets in a small number of locations, this makes the organization more vulnerable to successfully being attacked as the attacker has fewer sites on which to concentrate their attack.
- **Protecting DNS**
  Many IT organizations implement just two or three DNS servers. As such, DNS is an example of what was discussed in the preceding bullet – how IT organization are vulnerable because their key assets are located in a small number of locations.
- **Implementing robust, multi-tiered failover**
  Many IT organizations have implemented disaster recovery plans that call for there to be a stand-by data center that can support at least some of the organization's key

[8] DDoS-attacks-growing-in-size

applications if the primary data center fails. Distributing this functionality around a global network increases overall availability in general, and dramatically reduces the chance of an outage due to a DDoS attack in particular.

In order to be effective, a CBSS that provides Web application firewall functionality needs to be deployed as broadly as possible, preferably in tens of thousands of locations. When responding to an attack, the service must also be able to:

- Block or redirect requests based on characteristics such as the originating geographic location and whether or not the originating IP addresses are on either a whitelist or a blacklist.
- Direct traffic away from specific servers or regions under attack.
- Issue slow responses to the machines conducting the attack. The goal of this technique, known as tarpits[9], is to shut down the attacking machines while minimizing the impact on legitimate users.
- Direct the attack traffic back to the requesting machine at the DNS or HTTP level.

A CBSS that provides Web application firewall functionality is complimentary to a premise-based Web application firewall. That follows because while the Cloud-based Web application firewall service can perform many security functions that cannot be performed by an on premise Web application firewall, there are some security functions that are best performed by an on premise Web application firewall. An example of that is protecting an organization against information leakage by having an onsite Web application firewall perform deep packet inspection to detect if sensitive data such as a social security number or a credit card number is leaving the site. If sensitive data is leaving the site, the onsite Web application firewall, in conjunction with other security devices, can determine if that is authorized and if it is not, it can prevent the data from leaving the site.

## Evaluating the Security of Cloud Based Services

The primary concern that limits IT organization's use of public cloud services is concerns over security. Realizing that, the following is a set of security focused criteria that IT organizations can use to evaluate CCSP provided services.

- Can the CCSP pass the same security audits (e.g., PCI, HIPAA) to which the IT organization is subject?
- Does the CCSP undergo regular third party risk assessment audits and will the CCSP make the results of those audits available to both existing and potential customers?
- What are the encryption capabilities that the CCSP provides?
- To what degree does the CCSP follow well-established guidelines such as the Federal Information Security Management Act (FISMA) or National Institute of Science and Technology (NIST) guidelines?

[9] Wikipedia Tarpit(networking)

- Has the CCSP achieved SAS 70 Type II security certification?
- Is it possible for the IT organization to dictate in which countries their data will be stored?
- What tools and processes has the CCSP implemented to avoid unauthorized access to confidential data?
- Will the CCSP inform the IT organization when someone accesses their data?
- Does the CCSP have the right and/or intention to make use of the data provided to it by the IT organization; e.g., analyzing it to target potential customers or to identify market trends?
- What are the CCSP's policies and procedures relative to data recovery?
- What procedures does the CCSP have in place to avoid issues such as virus attacks, Cross-site scripting (XSS) and man in the middle intercepts?
- How well trained and certified is the CCSP's staff in security matters?

## About the Webtorials® Editorial/Analyst Division

The Webtorials® Editorial/Analyst Division, a joint venture of industry veterans Steven Taylor and Jim Metzler, is devoted to performing in-depth analysis and research in focused areas such as Metro Ethernet and MPLS, as well as in areas that cross the traditional functional boundaries of IT, such as Unified Communications and Application Delivery. The Editorial/Analyst Division's focus is on providing actionable insight through custom research with a forward looking viewpoint. Through reports that examine industry dynamics from both a demand and a supply perspective, the firm educates the marketplace both on emerging trends and the role that IT products, services and processes play in responding to those trends.

Jim Metzler has a broad background in the IT industry. This includes being a software engineer, an engineering manager for high-speed data services for a major network service provider, a product manager for network hardware, a network manager at two Fortune 500 companies, and the principal of a consulting organization. In addition, he has created software tools for designing customer networks for a major network service provider and directed and performed market research at a major industry analyst firm. Jim's current interests include cloud networking and application delivery.

For more information and for additional Webtorials® Editorial/Analyst Division products, please contact Jim Metzler at jim@webtorials.com or Steven Taylor at taylor@webtorials.com.

**Published by Webtorials Editorial/Analyst Division**
www.Webtorials.com

**Division Cofounders:**
Jim Metzler
*jim@webtorials.com*
Steven Taylor
*taylor@webtorials.com*

**Professional Opinions Disclaimer**

All information presented and opinions expressed in this publication represent the current opinions of the author(s) based on professional judgment and best available information at the time of the presentation. Consequently, the information is subject to change, and no liability for advice presented is assumed. Ultimate responsibility for choice of appropriate solutions remains with the reader.

Customer Driven Innovation

# AX Series Application Delivery and Advanced Server Load Balancing

## Flexibility to Solve Critical Business Challenges

A10 Networks was founded with a mission to be the leader in Application Networking. With the rapid speed of innovation allowed by advances in communication, customers choose A10 Networks to help their applications keep pace.

It is predicted that by 2020, there will be 31 billion devices and four billion people connected to the Internet (source: Intel). This massive and accelerating growth in network traffic is driving Application Networking momentum. As business critical applications continue to grow in number and complexity, intelligent tools are required for efficient performance.

We are only touching the surface for what is possible today, and it is certain that the need for intelligent Application Networking tools will only increase. Predicting this trend, A10 developed a new generation platform with the flexibility to solve critical business challenges for three key initiatives: Any App, Any Cloud and Any Size.

## Any App

### Web Scalability and Availability

Today's web servers are conduits for complex applications that require intelligence at every layer. If an application is slow or unavailable, or an Internet connection or server goes down, business productivity and profits are lost. A10's flexible Application Networking platforms give customers full control of their web, and any application environment, enabling scalability and availability for all mission-critical applications. In addition, partnerships and certifications with major vendors such as Microsoft, Oracle and VMware, enable rapid and predictable deployments.

### IPv4 Exhaustion and IPv6 Migration

Amid rapid network growth, a key challenge is to ensure that expansion can continue unabated for brand protection and uninterrupted business, avoiding costly IT fire drills. A10 delivers powerful, enterprise and carrier class IPv4/IPv6 solutions at attractive price points that will enable organizations to extend and preserve existing IPv4 investments and provide a clear path to IPv6 while enabling communication and connectivity between the two protocols, with many of the largest deployments worldwide.

## Any Cloud

### Enterprises, Web Giants, Service Providers

With over 2,000 customers across all verticals, including companies such as GE Healthcare, LinkedIn and Microsoft, A10 has focused expertise to service constantly evolving network requirements with a rapid return on investment (ROI). Customer benefit examples include the ability to deploy differentiated customer services, reduce costs through data center consolidation, increase efficiency with large traffic volumes, accelerate web speed to drive customer satisfaction and many more. A10's flexible platform addresses needs for any cloud today, and in the future.

### Multi-tenancy and Virtual Clustering

A10 delivers multi-tenancy through advanced high-performance Application Delivery Partitions, allowing customers to provide many services and applications to different groups on a single platform, with full network separation and without any hidden license costs. Any organization sharing the same infrastructure can greatly reduce Total Cost of Ownership (TCO) for Application Networking. Unique clustering technology extends unmatched scaling from millions to billions of connections as required.

## Any Size

### On-demand Virtual Appliances

A10 offers virtual appliances via hypervisor solutions as alternatives to its hardware platforms. With scale-as-you-grow options in numerous different sizes, A10's virtual machines can be rapidly deployed on commodity hardware, scaling up and down on-demand for changing traffic volumes and use cases.

### Scalable and Faster Appliances

At A10, performance is a path to data center efficiency, and not the end itself. With the industry's fastest Application Networking platforms in the most compact form factors, A10's performance delivers overall optimization, ensuring non-stop commerce and applications with lower operational costs. All features are included without licenses so that additional budgets are not needed for new features, allowing for rapid deployments without any license complexity, streamlining internal operations.

## Contact us

Contact us today to discuss how A10's AX Series Application Networking platforms can solve critical business challenges within your mission-critical IT infrastructure: for any app, any cloud or any size.

**Tel: +1 888 A10-6363 | inquire@a10networks.com | www.a10networks.com**

# Aryaka's WAN Optimization as-a-Service Brings a Bold New Direction to the Modern Distributed Enterprise

THE CLOUD has become the next logical step in the evolution of optimizing the enterprise wide area network (WAN) for today's global workforce.

WAN optimization is about improving the performance of business applications over WAN connections. This means matching the allocation of WAN resources to business needs and deploying the optimization techniques that deliver measurable business benefits. Since the WAN is the foundation of the globally connected enterprise, the performance of the WAN is critical to business success.

In the last decade, enterprises seeking to improve application performance across the WAN had little choice but to symmetrically deploy hardware-heavy WAN optimization controllers in data centers and remote locations, invest further in bandwidth, provision MPLS links or a combination of these. These dated solutions do not scale, create other problems and are beyond the affordable reach of 90 percent of the world's businesses. Enterprises suffer inasmuch as underperforming applications have a significant impact on a company's operational performance, including slower access to critical information and higher IT costs.

New cloud-based WAN optimization as-a-Service technology changes all that. This technology better addresses application performance problems caused by bandwidth constraints, latency or protocol limitations. WAN optimization as-a-Service dramatically improves response time of business-critical applications over WAN links and maximizes the return on investment in WAN bandwidth. Enterprises can ensure collaboration and avoid the need for costly, complicated hardware appliances or dedicated MPLS links.

**The "Cloud" Defined,**
**WAN Architecture Redefined**

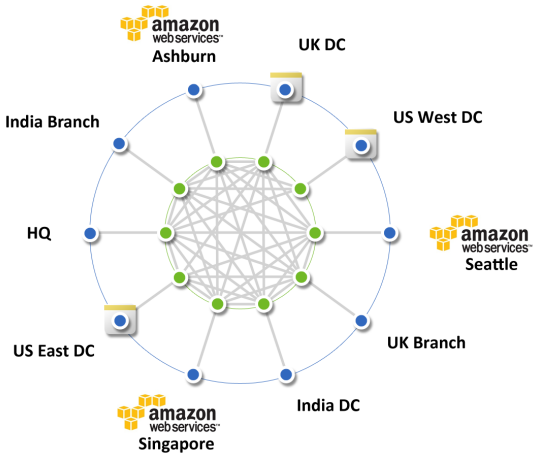
The term "cloud" is intriguing and varied in its description. Vendors within the WAN optimization space and other service providers are trying to find a way to optimize access to the cloud. The only way they can achieve this is by installing another appliance where possible – a virtual appliance – in limited situations within the cloud provider's infrastructure. The cloud for any enterprise can mean public, private or hybrid; it can be data or applications hosted within a private data center or offered as a global on-demand (SaaS) application. Every enterprise requiring optimized access to the cloud will have to install a virtual appliance for each cloud service they need to access, and another few at locations or users that want to access this cloud service.

There is a simpler way to achieve optimized access to cloud services worldwide, irrespective of their purpose and infrastructure location. Aryaka has created multiple Points of Presence (PoPs) across the world connected by a dedicated, secure and highly redundant network. This optimized network connects the enterprise WAN to any cloud service and remote locations worldwide in a simple, CAPEX-free, seamless way without any appliances or dedicated access links.

The cloud has redefined the architecture to optimize the enterprise WAN as the third and most important part needed for the success of compute and storage. Aryaka's purpose-built network drastically increases throughput to reduce the time required and data transmitted between enterprise locations and cloud services. Using compression, deduplication, Quality of Service (QoS) and TCP optimization technologies that are the cornerstones of these optimization solutions, enterprises can experience significant application performance gains 2-100X faster.

Global enterprises leveraging WAN optimization as-a-Service are improving productivity, enhancing collaboration and increasing network and application performance.

***"Simplicity is the ultimate sophistication."***
-Leonardo da Vinci




*An Aryaka customer's locations, data centers and Amazon instances are meshed to Aryaka's closest POPs to leverage transport of all traffic across one optimized network.*

Aryaka's WAN optimization as-a-Service solution is sophisticated simplicity. The solution eliminates the need for expensive and complex appliances as well as long-haul connectivity worldwide. With Aryaka's WAN optimization as-a-Service solution, globally distributed teams can communicate and collaborate with the security, reliability, end-to-end visibility and control required by the enterprise.

By SONAL PURI

## ABOUT ARYAKA

Aryaka is the world's first cloud-based WAN optimization as-a-Service company solving application and network performance issues faced by the distributed enterprise. Aryaka has been named to the Dow Jones VentureWire FASTech 50 innovative startups for 2011, a "Cool Vendor" by a leading analyst firm and a GigaOM Structure 50 company that will shape the future of cloud computing. Aryaka eliminates the need for expensive and complex WAN optimization appliances as well as long-haul connectivity, and enhances collaboration across corporate locations, data centers and cloud services. It offers significant cost benefits, ease-of-use, instant deployment, performance advantages, dramatic productivity gains and real-time insight into WAN applications, locations and performance while providing 24/7 world-class support. To learn more, visit www.aryaka.com. Follow us at Twitter, Facebook, YouTube and on LinkedIn.


Aryaka
691 S. Milpitas Blvd.
Milpitas, CA 95035
Tel: 1-877-727-9252
www.aryaka.com

# Optimize and Secure Cloud, SaaS, BYOD, and Social Media

How to Re-architect to Lower Networking Costs and Safely Improve Performance

So many of the dominant trends in applications and networking are driven from outside the organization, including cloud and Software-as-a-Service (SaaS), Bring Your Own Device (BYOD), Internet streaming video, and social networking. These technologies of an Internet connected world are fundamentally changing how we live and work every day. Yet, today's network and security architectures struggle to adapt.

A design that concentrates Internet access at a few data centers and backhauls branch Internet access over the Wide Area Network (WAN) is expensive; it creates overburdened networks and slows the response of both cloud-based and internally delivered applications. The reason this architecture persists is fear. Today's threat landscape has migrated to the web causing many security professionals to prevent direct Internet access at the branch.

But with new cloud-based security solutions from Blue Coat you can re-architect your network to embrace the Internet – safely – and optimize application performance.

## First: Re-Architect Branch Connectivity with Cloud-based Security to Lower Costs

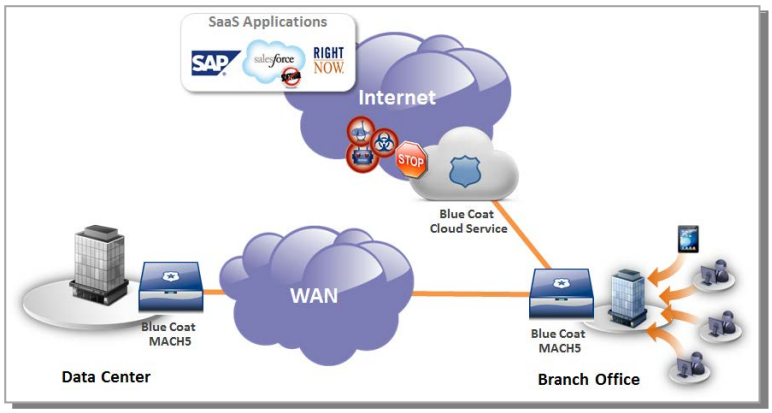
Blue Coat Cloud Service allows you to provide the same enterprise policies and technology to branch and mobile users. By leveraging Blue Coat WebPulse™, a collaborative defense powered by a global community of 75 million users, the Cloud Service is able to deliver real-time protection against the latest web threats from wherever users access the Internet.

WebPulse is based on sound analysis-system design principles:

- Massive input: WebPulse analyzes up to 1 billion web requests per day.
- In-depth analysis: 16 layers of analysis support over 80 categories in 55 languages.
- Granular policy: Up to 4 categories can be applied to each web request for multi-dimensional ratings.
- Speed: Automated systems process inputs – in most cases, in real time.
- Results: This collective intelligence allows WebPulse to block 3.3 million threats per day.

The Cloud Service extends WebPulse protection beyond the WAN, providing secure access to cloud and SaaS for all users at any location. The benefits are clear:

- Lower costs, better performance. By enabling branch Internet, you reduce Internet traffic on the WAN by 60-70%; and directly connected cloud users enjoy better performance.
- The Industry's best analysis and threat detection technology powered by WebPulse provide immediate, continuous protection against known and unknown web threats.
- Universal policy and reporting provides you a single pane of glass to configure policies and report on usage across your entire user base.




## Second: Optimize Performance

SaaS, BYOD, Video and Social Media present challenges to network capacity and user patience. Blue Coat WAN Optimization helps overcome these challenges.

Chatty protocols and multi-megabyte files can hurt SaaS performance. Video requirements destroy capacity plans. Blue Coat's asymmetric, on-demand video caching and live stream splitting boost video capacity up to 500x – whether it's corporate or recreational video. For SaaS, our CloudCaching Engine improves performance by 3-93x, dramatically raising productivity for SaaS users at branch locations.

And now Blue Coat MACH5 technology secures SaaS applications as it accelerates their performance. MACH5 connects directly to the Blue Coat Cloud Service, enforcing SaaS user policies and leveraging WebPulse to scan and filter cloud traffic. Branch users can access applications like SAP, Salesforce, and RightNow without the burden of bandwidth slowdowns or risk of malware threats.

# If this is you... We need to talk!

- ☐ Require maximum application performance
- ☐ Planning to move applications into a cloud
- ☐ Virtualizing your Applications and Storage
- ☐ Backups or replications don't complete overnight
- ☐ Need affordable acceleration for SOHO & remote users
- ☐ Need WAN Opp for any hardware platform or hypervisor

aCelera™

## Get the WAN Optimization solution with the "Strongest Virtualized Architecture" *

Download for yourself: info.certeon.com/certeon-marketplace/

Request a Demo: www.certeon.com/demo

**Certeon aCelera software - accelerated access for ANY User, ANY Application, ANY Network, ANY Device.**

**Deploy in any mix of hardware, virtualization platforms, storage technologies, networking equipment and service providers. Supporting any custom or off the shelf application.**

www.certeon.com | 781 425 5200 | 5 Wall Street, Burlington, MA 01803

certeon
Accelerate & Broaden Application Access

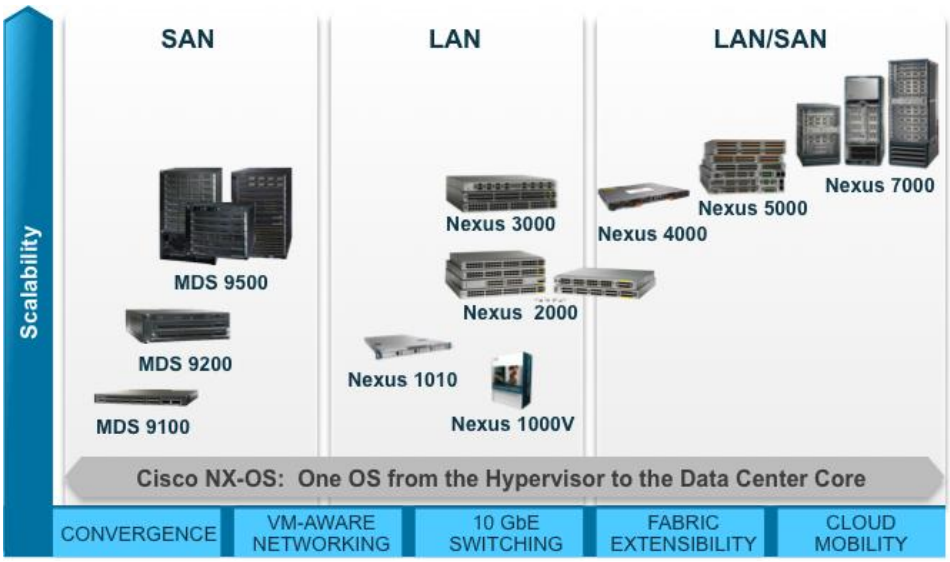

* Enterprise Management Associates

# Cisco Unified Fabric

Converged. Scalable. Intelligent.

Cisco Unified Fabric is a flexible, innovative, and proven platform for physical, virtual or cloud deployments. It provides the foundational connectivity within and across data centers so resources are highly available wherever and whenever they are needed.

A key building block for cloud-based environments and virtualized data centers, the Cisco Unified Fabric brings unmatched architectural flexibility and scale to meet the diverse requirements of massively scalable data centers, bare-metal infrastructures, high performance and big data applications.




- Revolutionary fabric scale with over twelve thousand 10 GbE server connectivity with Cisco Nexus
- Highest 10Gb Ethernet density in the industry with Cisco Nexus 7000
- High performance and ultra-low latency networking at scale with Cisco Nexus
- Network services delivered in virtual and physical form factors with Cisco ASA, ASA 1000v, WAAS, vWAAS, VSG and more
- Virtual networking from the hypervisor layer on up with Cisco Nexus 1000v, VSS, VDC, and more
- High availability within and across devices with ISSU, VSS, vPC, and more.
- Flattened and scalable networking at Layer 2 and Layer 3 with Cisco FabricPath, TRILL, L3 ECMP, and more
- Overcome the challenges of expanding networks across locations and the limitations of network segmentation at scale with Cisco OTV, LISP, VXLAN, and more
- Unified operational, control, and management paradigms across the entire fabric with Cisco NX-OS, DCNM and open APIs
- Converged networking to carry every kind of traffic on a single fabric with DCB and FCoE with Cisco Nexus and MDS

Cisco Unified Fabric is a flexible, innovative, and proven platform for physical, virtual or cloud deployments with a non-disruptive, evolutionary approach to create future-proofed, service- and cloud-ready data centers and prevent 'rip and replace' for existing data centers. For more info: http://www.cisco.com/go/unifiedfabric


**Americas Headquarters**
Cisco Systems, Inc.
San Jose, CA

**Asia Pacific Headquarters**
Cisco Systems (USA) Pte. Ltd.
Singapore

**Europe Headquarters**
Cisco Systems International BV
Amsterdam, The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at **www.cisco.com/go/offices.**


# Application Performance for Business Efficiency

*The unique way to guarantee business application performance over the WAN, increase IT productivity and save on IT costs.*

Ipanema Technologies – Fact Sheet 2012

## Business Overview

IT departments are witnessing change at a pace never seen before. Transformation is occurring as CIOs seek to access the benefits offered by unified communications, cloud computing, internet-based applications and consolidation, amongst many other strategic projects.

These initiatives are aimed at increasing an enterprise's business efficiency. While they simplify the way IT is delivered to users, they increase the complexity of corporate networking as applications and users rely on the continuous, reliable and consistent flow of data traffic.

Today many organizations are being held back from achieving the true value of their strategic IT programs due to overloaded and poorly understood networks, which were not designed for the symmetric, data-heavy, internet-driven environments that proliferate today. Application usage habits are changing rapidly too. Just a few years ago the extensive use of social media, video and unified communications applications was the exception. For many large enterprises it's now the norm. These new usages and applications have serious implications for the network. The change outlined above can have a dramatic impact, not least on the critical applications that support core functions of the business. Application performance problems including slowness and non- responsiveness impact the user experience and overall productivity of the organization.

In order to protect the business and the significant investments made in transformative applications such as unified communications and SaaS the network must be more intelligent, more responsive and more transparent.

## Ipanema at a Glance

- Corporate Headquarters: Paris (France)
- NA Headquarters: Waltham (MA)
- Used by worldwide market leaders across all industry sectors
- Over 150,000 managed sites with many 1,000+ site networks
- Leader for Application-Aware Network services
  (BT, Colt, C&WW, KDDI, KPN, OBS, Telecom Italia, Telefonica, Swisscom, etc.)
- Recognized as "Visionary" by Gartner
- A unique technology (Autonomic Networking) for automatic operations
- A system that tightly integrates all the necessary features
- A management platform that scales to over 400,000 sites

Ipanema automatically drives application performance over the enterprise's WAN from the priority of the business. With Ipanema, enterprises understand which applications run over their network, guarantee the performance they deliver to each user, succeed in their strategic IT transformations - like cloud computing, Unified Communications and hybrid networking - and control Internet traffic growth while reducing their IT expenses.

You can get Ipanema products through our distributor and reseller channels. You can also use them "as a Service" through numerous Managed Service Providers and Telecom Operators' offerings. SMBs/SMEs have access to Ipanema through AppsWork, a streamlined cloud service offering.

## Solution Overview

### Set your objectives and let Ipanema works for you – automatically!

Ipanema's revolutionary self-learning, self-managing and self-optimizing Autonomic Networking System™ (ANS) automatically manages all its tightly integrated features to guarantee the application performance your business requires over the global network:

- Global Application Visibility
- Per connection QoS and Control
- WAN Optimization
- Dynamic WAN Selection
- SLA-based Network Rightsizing

| Enterprise Applications | |
|---|---|
| Application | Criticality |
| SAP | Top |
| IP Telephony | Top |
| Telepresence | High |
| Logistics /Citrix | High |
| File sharing | Medium |
| Salesforce | Medium |
| Office 365 | Medium |
| SharePoint | Medium |
| Skype, Facebook | Low |
| YouTube | Low |

### Business efficiency requires guaranteed application performance

- Know which applications make use of your network…
- Guarantee the application performance you deliver to users…
- Manage cloud applications, Unified Communications and Internet growth at the same time…
- Do more with a smaller budget in a changing business environment, and to prove it…

### With Ipanema, control all your IT transformations

## For $3/user/month or less, you guarantee the performance of your business applications… and can save 10 times more!

Ipanema's global and integrated approach allows enterprises to align the application performance to their business requirements. With an average TCO of $3/employee/month, Ipanema directly saves x10 times more and protects investments that cost x100 times more:

- **Application performance assurance:** Companies invest an average of $300/employee/month to implement the applications that support their business. At a mere 1% of this cost, Ipanema can ensure they perform according their application SLAs in every circumstance, maximizing the users' productivity and customers' satisfaction. While they can be seen as "soft money", business efficiency and investment protection are real value to the enterprise.

- **Optimized IT efficiency:** Ipanema proactively prevents most of the application delivery performances problems that load the service desk. It automates change management and shortens the analysis of the remaining performance issues. Global KPIs simplify the implementation of WAN Governance and allow better decision making. This provides a very conservative direct saving of $15/employee/month.

- **Maximized network efficiency:** Ipanema's QoS & Control allows to at least doubling the actual capacity (goodput) of networks, deferring upgrades for several years and saving an average of $15/employee/month. Moreover, Ipanema enables hybrid networks to get access to large and inexpensive Internet resources without compromising the business, typically reducing the cost per Mbps by a factor of 3 to 5.

## What our customer say about us

***Do more with less***

"*Whilst data volume across the Global WAN has increased by 53%, network bandwidth upgrades have only grown by 6.3%. With Ipanema in place we have saved $987k this year alone*."

***Guarantee Unified Communications and increase network capacity***

*"Ipanema is protecting the performance our Unified Communication and Digital Signage applications, improving our efficiency as well as our customers' satisfaction. Moreover, we have been able to multiply our available capacity by 8 while preserving our budget at the same time."*

***Reduce costs in a cloud environment***

"*With Ipanema, we guaranteed the success of our cloud messaging and collaboration deployment in a hybrid network environment, while dividing per 3 the transfer cost of each gigabyte over our global network*."

## ABOUT IPANEMA TECHNOLOGIES

The Ipanema System enables any large enterprise to have full control and optimization of their global networks; private cloud, public cloud or both. It unifies performance across hybrid networks. It dynamically adapts to whatever is happening in the traffic and guarantees constant control of critical applications. It is the only system with a central management and reporting platform that scales to the levels required by Service Providers and large enterprises. With solutions used extensively by many of the world's largest telecom providers and enterprises across business and public sectors, Ipanema controls and optimizes over 100,000 sites among 1,000+ customers.

For more information www.ipanematech.com



www.ipanematech.com

# Do You Have the Best Choice in Application Delivery?

## Overview

The data center has some well known challenges - including application availability, performance and security – problems that can be addressed using Application Delivery Controllers (ADC). However, taking a closer look at businesses whose operations depend on agile and efficient data centers reveals additional challenges. Enterprise data centers need to scale flexibly in a cost-effective manner, ensure connectivity to current and next generation switching infrastructure, provide guaranteed reliability, be able to handle rapid growth and spikes in network traffic, and be capable of harnessing the benefits of virtualized resources and ecosystems. And of course, it goes without saying that all of these requirements must be satisfied while reducing both capital and operational expense.

Radware **Alteon® 5224** is an advanced ADC specifically targeted to address all of these challenges. Offering the very latest in next generation application delivery technology with benchmark affordability, it's simply the best application delivery choice.

Here are four reasons why, we know you'll appreciate:

## Reason 1: Unmatched OnDemand Scalability

The Alteon 5224 delivers unmatched on-demand scalability up to 16Gbps based on a simple software license-based mechanism. The platform supports the scaling of throughput capacity, additional advanced features and services (such as global server load balancing, bandwidth management, DoS protection and link optimization), as well as virtual ADC instances without device replacement or restart.

The result is that you pay only for the capacity you need. When you need more you upgrade the device you have and thereby eliminate costly capacity planning exercises and forklift upgrades projects. In contrast, if you were to scale from 1 to 16Gps with an ADC from a different vendor you may need to deploy up to 6 different platforms.

## Reason 2: Highest Performance in Class

Alteon 5224 offers the best all round performance metrics – compared to any other competing ADC platform in its class. It is simply the best solution for supporting traffic growth, can process more secured SSL transactions (for both 1024 and 2048 bit keys), and deliver more Connections per Second (CPS). All at the lowest price point available with:

- **3-8x more layer 4 CPS vs. F5** – delivering 500,000 layer 4 CPS
- **4-20x more layer 7 TPS vs. F5** – delivering 200,000 layer 7 TPS
- **1.5-3x more concurrent connections vs. F5** – delivering 12M concurrent connections
- **2.5-7x more SSL CPS (1024 bit keys) vs. F5** – delivering 35,000 SSL CPS
- **4-11x more SSL CPS (2048 bit keys) vs. F5** - delivering 11,200 SSL CPS

## Reason 3: The Only Enterprise Grade ADC with 10GE ports

Alteon 5224 is equipped with a total of 26 ports - the highest port density in the industry. This guarantees versatile connectivity options, enabling each Alteon 5224 to connect directly to more server farms or to ensure the physical separation of different networks without the need for intermediate switches. The result is simplified network architectures with fewer devices, reduced electrical and cooling costs, less rack space = greater savings.

In addition, Alteon 5224 offers a unique feature not found on any other 4Gbps ADC on the market: 10GE SFP+ ports. Connection to existing 1GE-interface switches as well as to next-generation 10GE-interface switches is straightforward. So as core switching fabric is refreshed over the next few years, the Alteon 5224 will continue to play well with its neighbors while your investment is protected.

## Reason 4: Virtualization Ready for Any Enterprise Size

Looking to virtualize your environment or already there? Alteon 5224 is capable of supporting multiple virtual ADCs on each physical device – each effectively equivalent in capabilities to a physical device.

How does it work? Similar to the concept of sever virtualization, each of the physical devices supplied as part of the Alteon 5224 can host a single ADC service or two ADC services or "instances" (at no additional charge) and can be expanded on-demand to support up to ten fully-independent vADC instances.

In addition, Alteon 5224 enables use of a separate vADC instance per application to ensure high application SLA compliance. The provisioning of additional vADC instances is easy and is achieved once again via on-demand software license updates with no service interruption. And all at a fraction of the cost of deploying additional hardware appliances.

## Simply Your Best Application Delivery Choice

The combination of these advantages – along with an industry unique 5-yeaar longevity guarantee – makes Alteon 5224 simply your best application delivery choice. Want to see for yourself? We invite you to download the competitive brief here or contact us at: info@radware.com.