nemertes RESEARCH INDEPENDENCE INTEGRITY INSIGHT — issue paper

Application Delivery for Virtualized Infrastructures

By Andreas M. Antonopoulos, SVP and Founding Partner, Nemertes Research and John E. Burke, Principal Research Analyst

Executive Summary

Virtualized infrastructures are highly dynamic and increasingly automated. By comparison, the network delivering applications to users tends still to be static, fragile, and manually configured. This mismatch can lead to erosion of the benefits of virtualization, errors, increased complexity, poor performance, and even downtime. The network needs application delivery optimization that can participate in the automated orchestration of virtual infrastructure to improve and guarantee performance and availability.

The Issue: Application Delivery in a Virtualizing Enterprise

Server virtualization is ubiquitous: 97% of organizations in Nemertes 2010-2011 benchmark use it.[1] On average, 49.4% of enterprise applications and 45.6% of mission critical applications now run on virtual servers. (Please see Figure 1.)

Although the desire to reduce capital costs and the need to survive within the space, power, and cooling limits of a data center drove most first-wave adoption of virtualization, agility has driven successive waves. Virtualization empowers rapid, inexpensive provisioning and de-provisioning of systems, letting IT respond more quickly to changing enterprise needs and letting enterprises try more new services, more quickly and for less expense. These agility benefits of virtualization are transformative: they don't just save a company money, they change the way a company does business, affect profitability and deliver competitive advantage through speed of execution.

A virtual enterprise seeks to free its staff as much as possible from the need to be in a specific place at a specific time in order to do business. IT is deploying applications including smart phones, virtual desktops, softphones, audio and video conferencing, and collaboration tools in support of enterprise virtualization. As agile IT uses data center virtualization to meet the needs of the virtual enterprise,

[1] All statistics drawn from Nemertes 2010-2011 benchmark. Nemertes benchmarks primarily organizations based in the U.S.; however, more than 51% of 2010-2011 benchmark participants have operations outside or are based outside North America.

tools like these will be in the forefront. One common characteristic among most of the new tools: an emphasis on real-time connectivity and a decreased tolerance for latency, packet loss, and other performance problems. And just as applications have become more demanding, users have too. They increasingly require access from anywhere at any time and have gradually lost tolerance for performance that varies with location.

In a virtualized infrastructure, dynamic resource management systems may provision and de-provision servers automatically to meet demand. They may locate them in data center with available capacity, or relocate them among racks or among data centers. Meeting performance targets in this environment is a tall order for most enterprises: the network is the least dynamic and evolved component in their virtualized infrastructures. In contrast to the new “speed of virtualization” enjoyed by servers and applications, the network used to deliver applications to end users in most enterprises remains as it has been for decades: manually configured, complex, fragile, and static. This mismatch can lead to erosion of the benefits of virtualization, increased complexity and accompanying errors, poor performance, and even downtime.

The Architecture: How Virtualization Changes Everything

Server virtualization, in abstracting logical server attributes from underlying hardware, creates three key differences between physical and virtual servers:

- The concentration of multiple servers onto one physical machine. Organizations host an average of 12 VMs per physical server, but some host 30 or more.

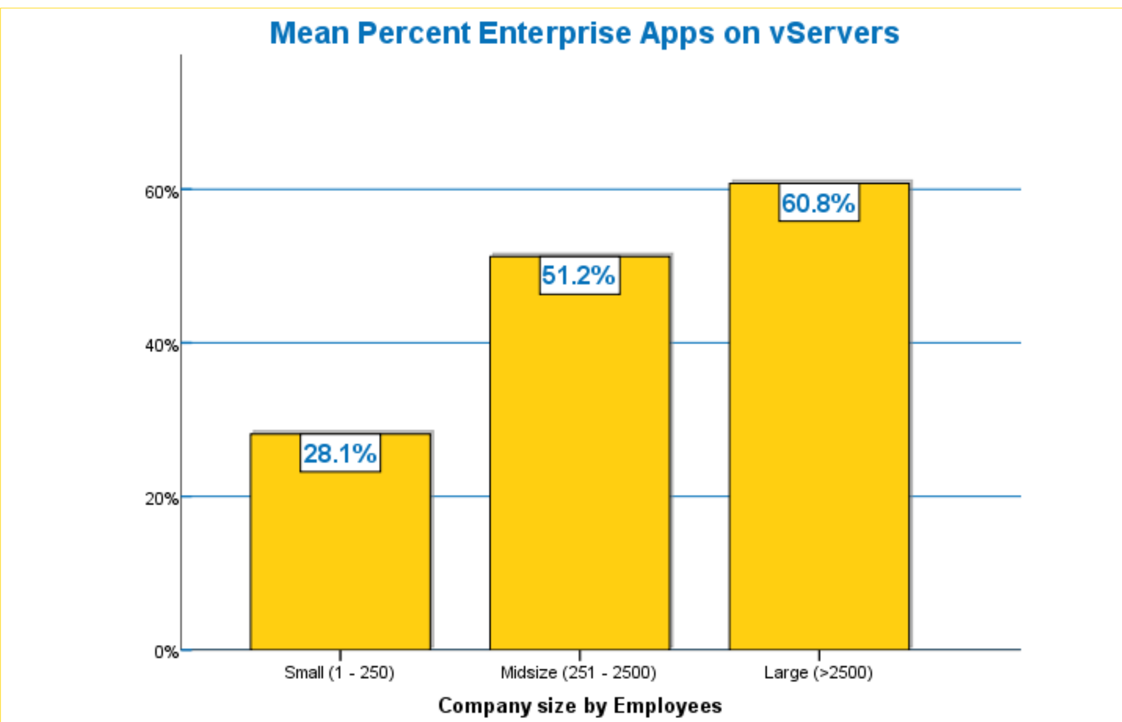

Figure 1: The Rise of Virtualized Application Stacks

- The variable state of servers over time: They can be active, frozen, re-activated, stored and reloaded.
- The movement of virtual machines among physical servers in the same data center, among data centers or even out to a cloud or hosting provider.

To address the challenges these facts create, enterprises need to adopt an application delivery optimization (ADO) strategy that explicitly addresses virtualization.

Application Delivery Optimization

Nemertes defines Application Delivery Optimization (ADO) as the holistic discipline of ensuring that applications are delivered with optimal performance, appropriate security, and continuous availability to an increasingly dispersed and distributed organization. ADO comprises a spectrum of interlocking technologies, including WAN optimization, security, and application acceleration. Options for implementing ADO range from appliances to software to services, and no one approach is right for all organizations. (Please see Figure 2.)

ADO for Virtual Enterprises

Application delivery optimization can act as the glue that holds applications and users together, and that joins dynamic resources to dynamic demand. The good news is that, properly deployed, ADO can deliver both technical benefits and business value to a virtualized environment.

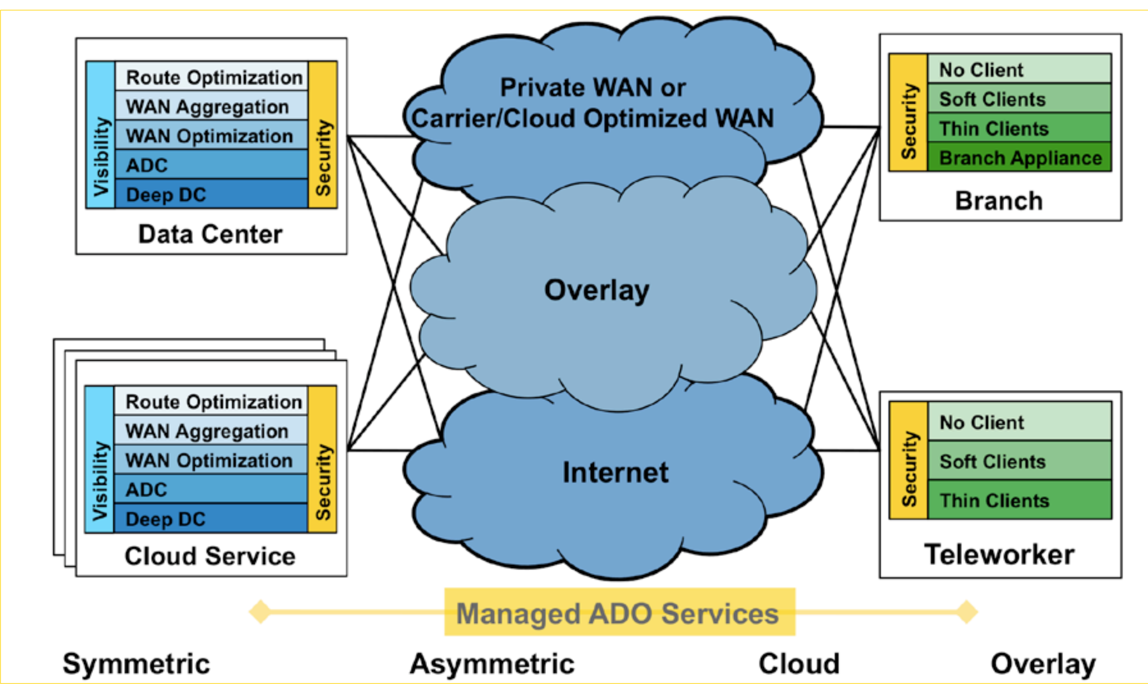

Figure 2: Application Delivery Optimization

Via application delivery controllers, global load balancers, and route optimizers, ADO can hide the complexity, mask the movement of servers from moving users, balance resource demand across available resources and so ensure IT can deliver the application reliably, responsively, securely and globally. Via compression, packet loss mitigation, and application accelerations ADO can improve performance for users on low-bandwidth connections.

Most organizations use some ADO technology, most typically in the form of application delivery controllers or WAN optimization controllers. More aggressive IT organizations, which deploy all virtual enterprise technologies more often than less aggressive IT organizations do, use ADO more often, too. They are expanding both the breadth and depth of their ADO adoption: more such organizations use ADO, and they use it for more locations. (Depth of deployment averages 43.1% of locations in more aggressive organizations compared to 37.3% among less aggressive organizations).

Application Delivery Optimization, Holistically

Companies say that one of the main challenges they face in virtualizing their data centers is adapting internal processes to the new, virtualized environment. The biggest factor: the blurring of lines of responsibility. In a traditional non-virtualized data center, teams manage different resources without much need for coordination and overlap. In a virtualized data center, IT shops must manage servers, storage, networking and security together, often in tight coordination and with blurred boundaries. Hypervisors contain virtual network switches, can re-map storage connections, and can move both across the data center. Applications are deployed as a collection of virtual machines, with storage and virtual networking services configured at the same time, often within the same tool. IT teams find that previously clear lines of demarcation between the resources they manage are now blurred.

In the virtualizing environment, networks are getting flatter and fatter, heading toward huge, consolidated, high-speed, layer-two switched networks combining virtual and physical switches in support of the virtual infrastructure. This shift is pushing data-center network architectures from three-tier (edge, aggregation, and core) to two-tier network (edge and core), squeezing out the aggregation layer.

Traditionally, companies have adopted various security and optimization features for application delivery only as needed, as band-aids. If an application performs poorly, the network manager may buy an acceleration appliance, or a caching appliance or a load balancer to solve the problem. When another application in the data center suffers from poor performance, the network manager will buy more optimization. The optimizers often get slotted into the same infrastructure silo hosting applications and their data.

This ad-hoc approach worked fine when most applications existed in separate and static silos and most users had physical desktops in a high-performance LAN. One-off solutions could fix problems, at least for a time. Today, because the problem of application delivery has become more complex this approach is neither sufficient nor cost effective. Their virtualized servers (and, increasingly, virtualized desktops) have to use decidedly un-virtualized, static networks that are not able to keep up with the speed of innovation. Where a server manager can provision a new application in a matter of minutes or hours, it might take weeks to get the network configured to secure, optimize, load balance, cache and offer seamless disaster recovery for that application.

Now, application delivery optimization is not just about fixing *a* performance problem between point A and point B, it is about bringing together fluid applications and mobile users across vast geographies with the highest level of performance and reliability. ADO has to solve a spectrum of problems that may change rapidly and repeatedly for servers and users that don't stay put.

So, organizations are building application delivery optimization into the data center pervasively, a feature of the architecture. It rises above silos where they still exist. As cloud-style rapid application deployment moves into a data center, it moves into the virtual infrastructure, ensuring that any application can be delivered with the same level of quality, efficiency and security as in the old-style, silo-per-app paradigm. Rather than applying multiple band-aids, network managers and application delivery architects design an ADO architecture that spans data centers and the WAN and addresses the performance problems of users inside the company net, on the VPN, or using the Internet.

Virtualizing ADO

The main ways to match ADO flexibility and speed to that of the new data center are to support multi-tenancy in ADO devices, and to virtualize appliances. Multi-tenancy, the ability to support separate domains of action within a single appliance, is critical in the modern data center. It supports easier asset utilization tracking, a key discipline of the service-management philosophy that is on the ascendant in most IT operations. It also enables separation of systems, key to meeting compliance requirements, mitigating risks, and improving network security. Multi-tenancy can be supported within a single operating instance of an appliance (for example, a single logical ADC that knows how to separate different groups of users and servers) or through in-chassis virtualization of ADO appliances (for example, a single hardware appliance hosting multiple, separate logical ADCs).

Virtualized ADO appliances packaged as standard virtual machines using either OVF or platform-specific standards can be handled like other virtual machines. They can be spun up at need, run as long as needed, spun down when the need passes. They can be installed within, moved with, and spun down at the same time as the virtualized application stacks they assist. They can even flow out onto external public or private cloud infrastructure, to optimize application traffic to and from servers there for as long as the servers persist.

Orchestration, Automation and Dynamic Scaling

Servers, groups of related servers, even entire virtual data centers, can be provisioned and de-provisioned quickly in a virtual environment. Automated dynamic scaling—having management systems that automatically spin up new units of application function as current and projected demand dictate—is a key requirement for fully realizing the enterprise benefits of virtualized infrastructure. Deploying it is a critical step in transforming a data center into an internal (private) cloud.

When the application environment becomes elastic in this way, the network and any ADO services within it must become similarly elastic. Close coordination of changes in back-end resources with the reconfiguration of network services between those resources and users is the only way to prevent ADO tools from becoming a bottleneck. When a layer of network services can't change as rapidly as the resources behind it, it becomes a brake on innovation that continuously thwarts the business's attempts to create a responsive, agile application environment.

ADO solutions need to share an orchestration environment with the other virtual resources. Where ADO is a layer outside the virtual infrastructure, ADO devices need to reconfigure in response to orchestration-system messages about changes in virtual servers, remapping rule sets or spinning virtual instances up and down to take into account the appearance or disappearance or movement of servers. Where ADO is provided using virtual appliances, obviously, these virtual machines need to be added to the application stacks they serve, for the orchestration system to spin up and down as a unit.

It is also important that the orchestration system listen to the optimization layer. ADO devices such as local and global load balancers can provide orchestration tools critical information for deciding when and where to spin up new virtual infrastructure as demand grows or shifts, and for deciding when the need for an application component or stack has passed.

Business Continuity and Disaster Recovery Considerations

As organizations move along the road to private cloud, trying to make their infrastructure agile and responsive, they also move from a "disaster recovery" thinking to a "business continuity" model. The shift to business continuity is a testament to the fact that virtualized infrastructure puts real continuity of service into the reach of far more organizations than before. Combining virtual infrastructure with virtual desktops, which make continuity of enterprise desktop access really possible, and with ADO to mitigate or mask transition to fail-over facilities, IT can deliver unprecedented levels of availability and reliability. The old model of data center reliability involved a "hot" data center with a "cold standby" ready to take over if disaster struck. With dynamic resource pools, global load balancers and optimization, it is now possible to make every data center a production "hot" data center. Application delivery controllers and route optimizers can push users to the closest data center, ensuring efficient use of resources when all is well, and continuity of access when disaster strikes. Localized problems (like losing a data center) become another a simple resource constraint, with the ADO layer redirecting users to locations where resources are still available and letting the orchestration systems know that demand will be ramping up rapidly.

Security Considerations

As noted, supporting security is a critical component of an ADO architecture. It can do so in a few ways: helping ensure only appropriate, authorized access to applications being delivered; helping make sure that only desirable traffic is optimized for delivery; and helping protect data in transit.

ADO devices often sit at critical chokepoints in the WAN and on Internet connections. They should therefore be able participate in identity-driven security, for example by integrating with VPN gateways or network access controllers to prevent network (and so) application access by unauthorized users.

Likewise, they can participate in traffic filtering, performing basic (or advanced) URL blocking, malware removal, and worm or virus blocking. By helping eliminate bad traffic, they can further optimize networks for good traffic.

Lastly, and most specific to a virtualized infrastructure, ADO systems can improve virtual security. The dynamics of virtualization violate a fundamental assumption of the traditional "strong perimeter" defense: that servers and applications don't move. Under virtualization, they do, in fact move, sometimes outside the perimeter (for example, to the cloud or an external data center or a branch). ADO devices can not only speed this movement among host sites; they can add security, for example via added encryption.

Conclusions and Recommendations

Virtualizing enterprises require technologies to make access to resources independent of place and time. Agile enterprises require low overhead, highly dynamic IT. IT deploys virtualized infrastructures to meet the needs of the agile, virtualizing enterprise. Virtualized data centers are highly dynamic and increasingly automated; automation is the only way to reliably deploy virtual infrastructure on-demand, and to swiftly, flexibly de-allocate and reallocate resources.

The enterprise network, tasked with delivering applications to users, by contrast tends still to be what it has been for decades: manually configured, and therefore slower and more difficult to change without introducing configuration errors and increased complexity, leading to poor performance and even downtime. The relative slowness of network change acts as a brake on agile delivery, and erodes the benefits of virtualization.

The network needs application delivery optimization. ADO can mask from users the dynamism of the infrastructure and mitigate performance problems in a work-anywhere world. Organizations working toward implementing fully dynamic data centers and private clouds should be actively seeking ADO solutions that can support their goals, via multi-tenancy or virtualization or both. That optimization layer has to fully participate in the virtual infrastructure's automated orchestration to improve and guarantee performance and availability.

About Nemertes Research: Nemertes Research is a research-advisory firm that specializes in analyzing and quantifying the business value of emerging technologies. You can learn more about Nemertes Research at our Website, www.nemertes.com, or contact us directly at research@nemertes.com.