ISSUE PAPER

Securing Virtualized Infrastructure:

From Static Security to Virtual Shields

By Andreas Antonopoulos, SVP & Founding Partner, Nemertes Research

Executive Summary

Data centers have undergone tremendous transformation in the past five years. Virtualization has changed the architecture for servers, storage and networks. IT organizations are using management and provisioning automation to reduce operational cost, and increase responsiveness to business demands. But security is becoming more and more challenging. While virtualization has impacted every part of IT, security lags. As a result, most security products today do not easily support a virtualized infrastructure. Next generation security products are emerging however. Implemented as virtual software appliances or as virtual shields, they can provide security in a pool of virtual servers, regardless of operating system or application. Data center architects and security architects need to carefully choose security products that support -- and not subvert -- their data center virtualization strategies.

The Issue: A New World to Secure

Data centers today are truly "new" from every perspective: facilities, storage, management, computing, and networking. Although data centers have existed as long as enterprise computing itself has, a confluence of economic, enterprise, and technological changes is driving a major metamorphosis in data center design and implementation. This, in turn, is determining how data center and security professionals approach the problem of securing the data center and the enterprise network from threats, internal and external.

Server Virtualization

Server virtualization is one of the most discussed technologies of 2005 and 2006. Despite all the hype, we find many organizations are not only taking a serious look at server virtualization, but they also are

generating substantial savings through increased utilization, purchasing postponement, standardization and management automation.

Virtualization software vendors offer some interesting management tools that augment the hypervisor software. One class of such tools is the "live-migration" tool, first implemented by VMWare in the VMotion technology. Other vendors, including XenSource, Virtual Iron, SWSoft and Microsoft, are reportedly working on similar tools allowing the live migration of virtual machines.

Live migration is a technology that can move a virtual machine to a different physical host, without any impact to end users. Effectively, the virtual machine stores its state onto shared storage, just prior to migration. On demand, the virtual machine is re-instantiated with its virtual-network connections and memory state intact, and can continue to serve users with no noticeable interruption. This type of solution is especially useful given the extreme availability demands we have documented in Nemertes' research with data center managers. When end-users demand 100% availability, one of the most common complaints we hear from IT managers is the loss of the "maintenance window" (ie, the ability to have planned downtime for services in order to perform maintenance on hardware or software). With live migration, administrators can move virtual machines moved off a server, without disruption, power down the server or reboot it for maintenance, and migrate the virtual machines back once they finish maintenance (See Figure 1: "Virtualization For Continuity," on page 2).

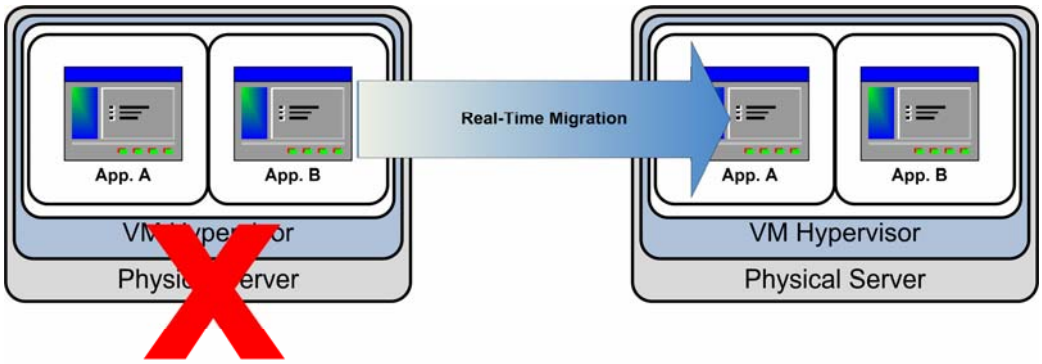

Figure 1: Virtualization For Continuity

Another use of live-migration technology is load balancing. If demand for a particular service starts increasing rapidly, IT operations engineers can migrate high-demand virtual servers to another physical server with more CPU resources (with either fewer or no other virtual machines competing for the resources).

So, in addition to server consolidation, server virtualization is adopted for many other reasons. With features like live-migration it is possible to create a tremendously flexible, responsive, recoverable and cost-effective infrastructure. By orchestrating servers, storage and networking resources, virtualization tools and management platforms can deliver some of the benefits of on-demand computing today.

But, with all these benefits come increasing security challenges. Virtual machines are more difficult to secure, especially with today's static security devices. In order to achieve the most return on investment with server virtualization, an IT manager would have to pool servers into one large pool. This offers tremendous flexibility in workload allocation. But it also has the effect of flattening an infrastructure.

Previously, an enterprise might have had several DMZ-like zones of firewalls separating a Web server from an application server from a back-end database. Now, is that enterprise comfortable with all those devices floating in one pool of servers? Then again, the IT department could create several pools separated by physical firewalls or segment everything with VLANs, but it would lose the flexibility of server pooling. These and other challenges with virtualization mean that we may have to re-envision security in a virtualized world. While the "v-word" (virtualization) has appeared next to almost every aspect of IT (servers, CPUs, storage, networks, software applications etc), it is conspicuously absent in the security field. With the exception of virtual partitions on firewalls, which is more of a multi-tenant feature, little is said in the security industry about virtualization and virtualized infrastructures.

Virtual Server Patching

When Nemertes conducted research on enterprise security, we found server patching is a big drain on IT staff resources. Ironically, because patches might cause instability in servers, IT staffs were least likely to regularly and quickly patch the most critical servers.

If these servers run in a virtual machine, however, then they can test patches on a non-production instance. Thus, they will discover any instabilities, conflicts, or bugs before putting the patches into production. In order for this solution to be applied in an enterprise, organizations require some formal change-management processes to keep track of different versions of the server images on the SAN.

As security professionals apply each patch, the virtual-server image needs to be re-captured and stored in the freeze-dried version on the SAN. Then once the testing process has verified that there are no problems with the patch, administrators can restart the production server from the new image. Of course, this process still requires a reboot of the production server in order to instantiate the new image. In some production environments a reboot would be unacceptable down-time, but this solution is still far superior to the risk of un-patched and vulnerable production server or the Russian-roulette style patching directly onto production.

One of the most important security threats for servers is malicious software introduced within the operating-system files. Operating systems are vulnerable to software that can infiltrate the system drivers and other low-level system libraries because of lax separation of user space and kernel space. Very often, key software utilities and programs will run with privileges that let them modify core files or processes. This lets malicious software such as key-loggers, Trojan horses, or even entire remote-management rootkits infiltrate the lowest layers of the operating system. Unfortunately, since the security software running on the operating system is at the kernel- or user-privilege level, malicious software potentially can disable the security software or install into a part of the operating system that is invisible to the security software.

One virtualization solution that could help with this problem is to locate security software outside the operating system and within or below the hypervisor. This security layer would not be accessible by the operating system and therefore would be much less vulnerable to infiltration. Furthermore, this security layer would lie at a privileged level in relation to the operating system and could monitor memory, processes, and file systems from the outside. Such monitoring would be invisible to and "untouchable" by the operating system and any unauthorized software running on the host.

One such example is a software-based firewall and intrusion-prevention system that sits between the network device and the virtual-network interface of the hypervisor. Any traffic into and out of the virtual machine could therefore be inspected and, if necessary, blocked by the security subsystem. Thus organizations get the security of a completely separate black-box device without the cost of a security appliance.

Hard on the Outside, Soft in the Middle

Defense that focuses completely on the traditional perimeter of the enterprise – the Internet link – is now known as "tootsie pop" security: hard on the outside, soft on the inside. That is, once the hardened outer shell was breached, the inside presented no problems whatsoever to further compromises. For a time, this was a livable security posture, but the time has passed primarily because of two things. The first is the increasing sophistication of attacks, which, as they move up the protocol stack to attack via SQL, HTTP, XML, or SIP, or to mix modes of attack, render the hardening of the outer enterprise perimeter to IP-level attacks less relevant. The second is the erosion of the perimeter itself, as workers increasingly seek to work anywhere and everywhere they might be, using whatever network is ready to hand, and to be able to use laptops that way but then hook them up to corporate nets on return to the office, while at the same time the enterprise must be quite open to the traffic of partners, suppliers and customers.

The key peril is the attack from inside, whether via a compromised server, desktop or laptop, via a compromised partner or supplier, or in the form of an actual attack by enterprise staff or contractors.

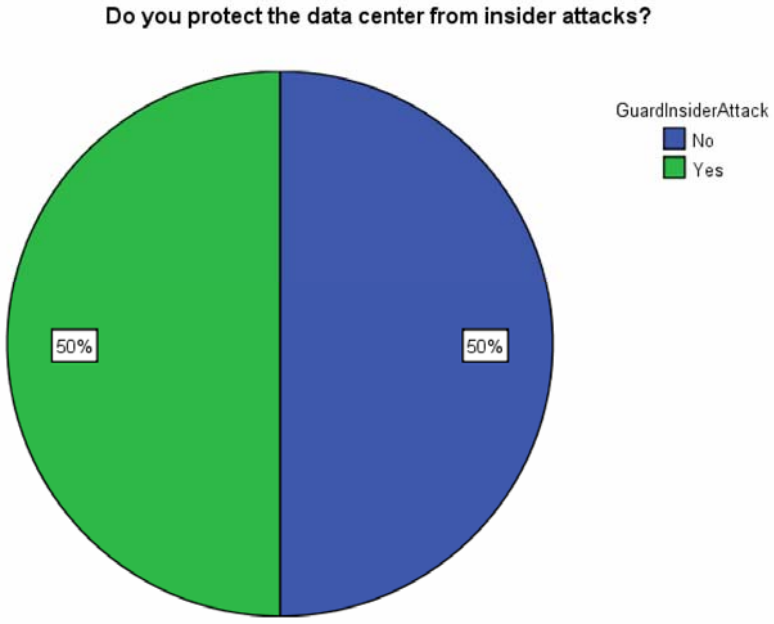

Figure 2: Defense Against the LAN

When we asked participants about their internal defenses, their responses indicated definite awareness and response to the insider threat, although at varying levels.

When asked whether they specifically set up defenses between their data centers and the enterprise LAN, only half said "yes." Many of the participants who were not doing so intended to deal with the problem mainly through improving security on endpoints; others had significant concerns about performance issues such as throughput and latency.

Once a server or other data center device is compromised, fast propagating attacks such as worms or viruses can quickly bring a data center to its knees, as many companies have discovered. Protecting the data center from such threats is not as simple as erecting a perimeter around it. One strategy to mitigate this type of risk is to compartmentalize the company's server plant– to build internal barriers between servers or groups of servers that can contain malware and stop it from spreading out of control.

When asked whether they took extra steps within the data center to protect servers from each other, to try to limit the damage caused by any compromise, a solid majority – a full two-thirds – said "yes."

(Please see Figure 3, "Data Center Security Segmentation," p. 6.) It is not surprising that a third of data centers do not pursue a segmentation strategy, as it can be expensive in both equipment and operational terms.

Security vs. Agility

The biggest operational drawback to this type of network segmentation is that it draws artificial lines through the company's infrastructure, leading to a more rigid architecture. The more entangled systems become, though, the more difficult such segmentation becomes. A firewall, for example, may be configured to allow access to a back-end database only from an authorized application server. Inside the firewall this will be represented by a rule containing the IP addresses of the database server and the application server. This type of static association makes a lot of sense in a world where resources are static, but is an entirely orthogonal model to the next-generation flexible data center. If the application has been automatically relocated to a different server and has a different IP address, the firewall will remain unaware and will break the connection with the database.

Many companies are implementing a business strategy focused on "agility" - ensuring they can respond to changes in their markets and deploy new applications rapidly. If the network is an enabler for new applications and rapid business change, static barriers will make it less flexible and therefore make the business less flexible also. This is especially so when virtualization and service-oriented architectures are taken into consideration.

If data center managers deploy security devices without considering the implications for their data center strategies, they may find that each security device deployed takes them one step further away from the next-generation data center. Flexibility in the data center through virtualization is a critical competitive advantage which should not be sacrificed because of security. Instead, data center managers should work very closely with security managers to help them understand the goals of the next-generation data center strategy, ensuring that any security purchases will be synergistic and not a setback. As data center technologies keep advancing towards virtualization, security vendors must pay attention and adapt to such an environment. After all, thwarting a data center strategy and sacrificing a company's competitiveness because of security is like throwing the baby out with the bath water.

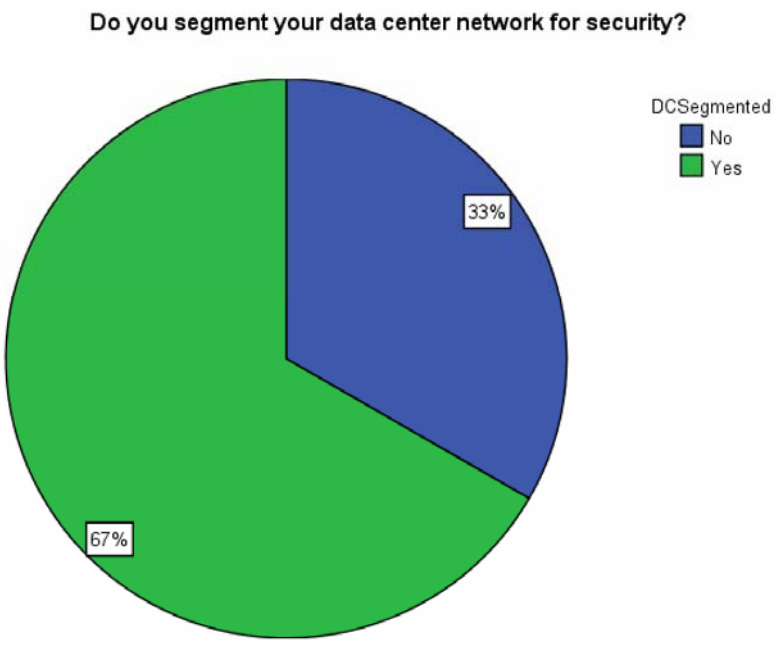

Figure 3: Data Center Security Segmentation

The data center is already hard to segment because it of the myriad interdependencies among systems. In a virtualized data center, hard servers become part of a resource pool and virtual servers can move among them – where would one put the security? How does one protect virtual servers?

A possible solution to segmentation inside the data center is to virtualize security. Instead of static security companies can build a virtual security layer to protect virtual systems. For a pool of servers in the data center, this type of security could be implemented in a virtual shield embedded in the virtual switch between the virtual servers and/or as a hypervisor plug-in.

Virtual Shields instead of Static Walls

What is the key characteristic of a virtualized data center? It's not about cramming more and more virtual servers into one physical server. It's about flexibility: rapidly deploying an application by orchestrating the necessary resources such as servers, storage, networking and security. Since the resources are abstracted from the underlying infrastructure, they can be provisioned with a lot of flexibility. Servers can be deployed as-needed from a pool, storage can be extended, networks can be partitioned with VLANs. In a virtualized environment all of these resources should be coordinated through a single management interface, as is the case with technologies such as VMWare's Virtual Infrastructure and Virtual Center. But what about security?

In today's data center, most of the security is provided by special purpose appliances. Contrary to the general trend of virtualization, these appliances are all-too-tangible. If a data center manager was to arrange all the servers in a big pool, the security devices would, by necessity, form a static ring around the pool. Segmentation of some sort is necessary, since the server pool may contain servers of different "tiers" of criticality (for example, Web servers facing the public and databases containing sensitive data). The servers need to be protected from the outside world, but they also need to be protected from *each other*. If a

single server in the pool is infected with a rapidly propagating threat then it will be able to cross-infect all other servers that contain the same exposed vulnerability.

Up to now, data center, security and network architects have had to reach a compromise of sorts – servers are on separate virtual networks (VLANs), which are switched through firewalls sitting in a ring around the server pool. Effectively, we end up using network virtualization to compensate for the lack of security virtualization. This approach is far from ideal, however. Since the security devices are static, they cannot respond to changes in the virtual servers. Let's say for example that a virtual server has to be moved to another physical server for maintenance. The security associations have to follow that server, so in order to keep things working, the server must retain the same IP address and VLAN as before.

Because of limited orchestration between the virtual servers and the non-virtual security, everything has to be done with VLANs. The disadvantage is that VLANs are difficult to manage and they are too coarse-grained for use as security controls. If you bunch all databases together you end up with the risk of cross-contamination. If you split each server into a separate VLAN you run out of VLANs. And in either case you have a management mess on your hands.

Part of the conundrum arises from some assumptions inherent in the design of almost all security devices. Firstly, there is an assumption that network address is the same as server identity. All access control lists for example are based on the same model: IP address or subnet **"A"** is allowed access to IP address or subnet **"B"**. This works well when servers are static and fixed. But once you add mobility and virtualization, you start revealing the faulty assumptions in that model. IP address A *was* the Web server, but it has since been moved to another subnet, or another IP address through re-provisioning or live-migration. Did anyone remember to tell the firewall? Network addresses are a *proxy* for the identity of the server or application that is in fact independent of it's location on the network, and this proxy relationship becomes weaker as server technology changes. Another fundamental design assumption is contained in the word "firewall" itself. The "wall" component refers to an immovable barrier, a segmentation device that is "architecturally" fixed and solid. In the fluid world of on-demand computing and virtualization, an immovable object can only cause one thing: turbulence. Furthermore, the "wall" is part of a broader paradigm of the "perimeter": large zones containing similar systems that can all be "trusted." With the level of interdependency between systems increasing and the level of interconnectedness with the outside world also increasing, the perimeter model is fundamentally flawed. Why do modern cities not have walls around them? Because cities depend on an intricate global network of trade, with fluid movement of people and goods through every part of the city.

In a modern city, security is not implemented through barriers and walls, not unless it is a war zone. Unless you want to sacrifice the business by taking the posture of a war zone, you have to apply security in a flexible and distributed manner. In a modern city, police roam around and respond to threats flexibly. A more apt metaphor for this kind of security is a *shield* in place of a wall. A *shield* is a light, portable and dynamic defense mechanism. It can be brought to the point of attack and does not wait for the attack to reach it. Most importantly however it can be carried, defending without limiting mobility. Bringing the *shield* concept back to the new data center model, we need a flexible, dynamic and virtual security device that can move with the servers it protects and be "orchestrated" by a management system on-demand. When we need to provision a new application, we allocate the server, storage, network *and security* resources dynamically.

Virtualized Security Models

While there are few products in the virtualized security space, some vendors such as Blue Lane, Reflex Security and StillSecure are beginning to create virtualized security solutions. What form would solutions in this space take? There are a number of architectures that we would expect to see:

- Security inside the hypervisor – The hypervisor itself can contain security code provided by the vendor, eg. VMWare. This would allow the hypervisor to provide security by default to all virtual machines. Hypervisors already contain some basic security functionality. However, we do not expect to see rapid development of security features in the hypervisor as it detracts from the core focus of the hypervisor vendor.
- Virtual Appliances - Based on server virtualization technology, a virtual security appliance is a software appliance running inside a virtual machine. Many software-based security solutions can be converted quite easily into virtual appliances. Some examples can be found in the VMWare Virtual Appliance Marketplace. Virtual appliances have the benefits of an appliance such as turn-key operation with the added advantages of virtualization such as live-migration and on-demand deployment.
- Security as a plug-in to the hypervisor – Security software could operate as a plug-in to the hypervisor. Using the API of the hypervisor, this security "layer" could intercept the network I/O stream, memory I/O stream or even CPU I/O stream. From this position, virtualized security software can monitor and modify data while remaining completely outside the operating system, opaque and isolated. Security services could include IDS/IPS, firewall, inline patch, cryptographic acceleration, in-line encryption, vulnerability scanning and integrity checking. Furthermore, we envision possible interaction with embedded security chips such as the Trusted Platform Module (TPM) in most new servers.

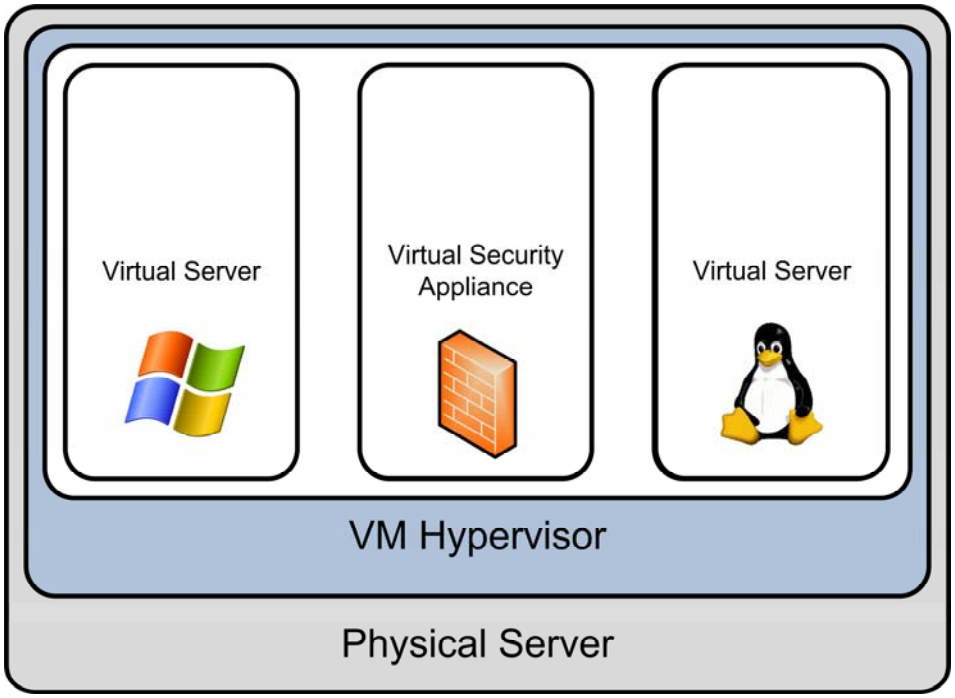

Figure 4: Virtual Security as a Virtual Appliance

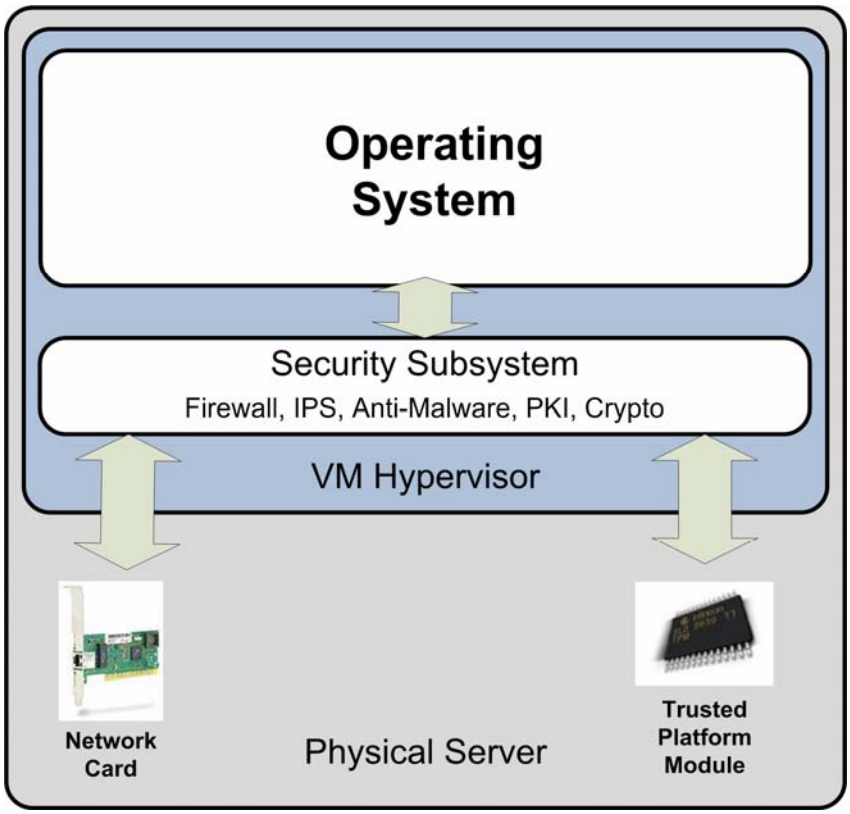

Figure 5: Virtual Security As A Plug-In

Conclusions and Recommendations

Business demands are driving tremendous transformation in the data center. Servers, storage and networking technologies have been transformed through virtualization. Meanwhile however, security virtualization has been lagging. As a result it has become more and more challenging for security architects to support these new virtualized data center architectures with static security devices. A newly emerging security paradigm, more akin to "shields" than "walls," consists of virtual security appliances that can be dynamically provisioned along side virtual servers. This new security model allows each virtual server in a pool to be protected both from the outside world and from other servers in the pool. As data centers continue to move towards on-demand computing through much broader adoption of virtualization, security must keep pace. Security architects need to carefully select products that are supportive of virtualized infrastructures. After all you would not want your security strategy undermining your data center and business strategy.

About Nemertes Research:

Founded in 2002, Nemertes Research specializes in analyzing the business value of emerging technologies for IT executives, vendors, and venture capitalists. Recent and upcoming research includes Web services, security, IP telephony, collaboration technologies, and bandwidth optimization.

About Analysts:
Andreas M. Antonopoulos, Senior Vice President & Founding Partner, is a renowned expert in security, networking and data-center technologies.

For more information, please contact research@nemertes.com.