



White Paper

metagroup.com



800-945-META [6382]

May 2004

The Evolution of Network Security: *From DMZ Designs to Devices*

A META Group White Paper

“Increasingly sophisticated cyberattacks, business and economic requirements, and even regulatory considerations are driving changes to both network security devices and their physical and logical arrangement. To ensure effectiveness, efficiency, and flexibility, solutions must evolve inward, upward, and outward.”



METAGROUP

Contents

Introduction	2
The Evolution of Network Security Designs	2
<i>Perfecting the "Perimeter": Long Live the Internet DMZ!</i>	2
<i>Extending the "Perimeter": Achieving Internal Isolation</i>	5
Business-Unit Barriers	5
The Data Center DMZ	6
The Evolution of Network Security Devices/Technology	7
<i>Strengthening the "Perimeter"</i>	7
Application-Layer Awareness and Control	7
Enhanced Attack Protection	9
Multiservice Security Gateways	10
<i>Extending the "Perimeter": Outward Evolution</i>	12
The SSL VPN Revolution	12
Bottom Line	14

Introduction

Despite significant investments in information security, organizations continue to be afflicted by cyber-incidents. At the same time, executive management is demanding that greater results be achieved with fewer resources. In other words, improving security effectiveness remains necessary, if not imperative, while enhancement of both efficiency and flexibility has also become a leading objective.

This paper explores how network security solutions must evolve to meet these requirements. Specifically, it examines changes required of both high-level network security designs (i.e., implementations, or layers) and the individual components (i.e., devices or products) that comprise them. Related topics include: evolving Internet demilitarized zones (DMZs), establishing internal DMZs, what “deep packet inspection” really means, the role of all-in-one security gateways, and the adoption and ultimate domination of Secure Sockets Layer virtual private networks (SSL VPNs).

The Evolution of Network Security Designs

We contend that achieving effectiveness, efficiency, and flexibility is dependent to no small extent on employing appropriate and comprehensive designs that are inherently tuned to these objectives.

Perfecting the “Perimeter”: Long Live the Internet DMZ!

Internet demilitarized zones (DMZs) have long been considered synonymous with network security (see Figure 1). This is unfortunate because, although they are necessary, they are also insufficient. More relevant to this section, though, is the fact that Internet DMZs must evolve to keep pace with changing requirements.

Figure 1 — Demilitarized Zones

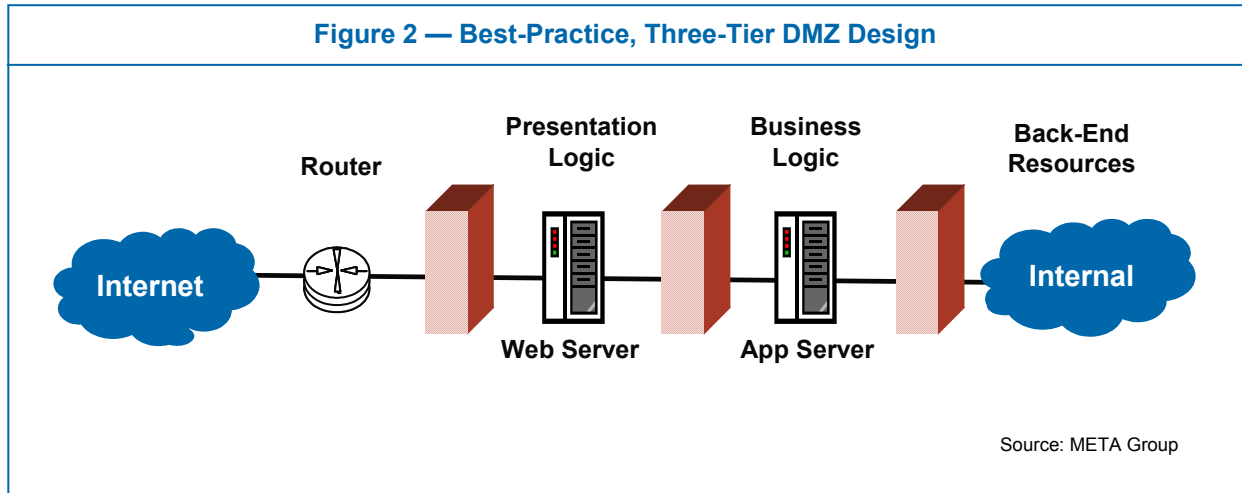
DMZs are implementations of firewalls and other security controls in an arrangement that is intended to form a buffer or transition environment between networks with different trust levels. For example, an Internet DMZ isolates an organization’s sensitive, internal computing resources from the Internet (an untrusted networking environment). Intermediate zones within a DMZ can also be used to host computing services, thereby making them accessible to external parties.

Source: META Group

Over time, a best-practice design that has emerged is the three-tier option depicted in Figure 2. This design has become popular because it affords the opportunity to apply different security filters and services between each tier of common three-tier Web applications, thereby thoroughly facilitating a defense-in-depth strategy.

However, although this design is still adequate, it is not ideally suited to today's objectives, such as minimizing the amount of exposed infrastructure (i.e., Web servers) and economically accommodating access by both internal and external users to the same set of applications.

Figure 2 — Best-Practice, Three-Tier DMZ Design



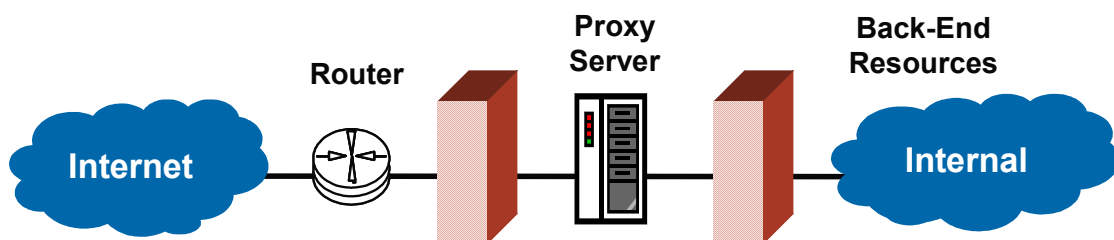
Therefore, we expect to see many organizations migrate to the two-tier, proxy-enabled design depicted in Figure 3 below. One very significant characteristic of this design is that the Web and application servers are no longer “visible,” having become part of the “back-end resources” (which also include integration and database servers). Overall, this approach has numerous advantages, including the following:

- **Enhanced security is derived:** This results in part from the design providing additional protection for the involved Web servers, which are notorious as weak points in any design. This protection stems from both the additional intervening firewall and the proxy server, which essentially acts as a new front end to the application environment. The proxy server not only enforces user authentication prior to allowing sessions to reach the Web servers, but also eliminates many network-layer attacks and affords the opportunity to apply even further filtering — by virtue of it terminating the original connection, providing an opportunity for detailed inspection, and then rebuilding the packets and connection with its own IP stack.
- **Improved economics and flexibility can be realized:** Fewer total servers may be required, since the proxy server can more readily support a one-to-many relationship with the application environment (potentially eliminating the need for some of the organization's Web servers). In addition, this arrangement can support access to the same applications by both external and internal parties without having to install separate instances of the application or having to resort to cumbersome and potentially insecure configurations (such

as routing internal users backwards through the DMZ, or routing them “out” to the Internet and then back in through the customer/partner-facing DMZ infrastructure).

- **The design is better aligned with the DMZ structure required for other computing services:** As a result, it has the potential to reduce the need for maintaining separate infrastructure for each service. In particular, we believe that Web services technology — with its component-to-component communications model and the likely need for a specialized security proxy — will be better accommodated by a two-tier DMZ design. In addition, a two-tier design is consistent with that used to support employee remote access and typical B2B connections.

Figure 3 — Two-Tier, Proxy-Enabled DMZ Design



Source: META Group

Another noticeable characteristic of this two-tier, proxy-enabled design is the apparent shortcoming that results from having one less firewall/security tier than the “best practice” design. This would be a legitimate concern, except for the fact that the two-tier design is incomplete as shown. Specifically, the intent is that this Internet DMZ design will be supplemented by implementing firewall/security tiers within the “back end,” or internal network.

However, before we elaborate on this relationship between the Internet DMZ and controls on the internal network, it is necessary to acknowledge that the DMZ designs depicted herein are not the only ones that we encounter or that we expect to see in the future. Numerous variations are inevitable and indeed necessary to account for the unique priorities and requirements of different organizations. Yet we do expect that the general characteristics, principles, and objectives enumerated above will remain the primary considerations behind whatever variation an organization eventually adopts.

In addition, it should be recognized that Figures 2 to 4 are not intended to be fully illustrative of all the details that comprise a real-world implementation. For example, Figure 3 should not be interpreted as requiring that all servers for externally accessible applications must be located within the internal environment. Indeed, we fully expect a wide variety of services will reside in the same logical tier as the front-end proxy server (e.g., e-mail relay/gateways, external domain name services, remote access concentrators, publishing/information-only Web servers). On the other hand, high-value transactional applications will most likely be deployed to take advantage of the front-end proxy architecture, and will in fact have most of their computing resources located within the internal environment.

Extending the “Perimeter”: Achieving Internal Isolation

The days of having a single choke point through which to control user access and inspect network traffic are long gone. Business realities demand support for mobile users, business partners, guests, and transient users (i.e., employees who operate alternately “from the field” and from within the office). The myriad connections into the corporate computing environment required by these users complicate, if not fracture entirely, the traditional perimeter-focused security designs. Furthermore, it is also necessary to account for internal users, who increasingly are being acknowledged as a significant source of threat in their own right — regardless of whether they are intentionally being malicious or are merely unknowing pawns in a larger battle being waged by an aggressive third party.

The result is the need to extend conventional network security controls inward. Specifically, internal “perimeters” should be implemented to provide containment and isolation services that are more localized to critical resources on the internal network. To be clear, the intent is not to advocate dissolution of the Internet DMZ — though a change or two will definitely be warranted, as already discussed. Rather the intent is to supplement this insufficient measure by judiciously implementing business-unit and/or data center DMZ constructs.

Business-Unit Barriers

One approach to achieving internal isolation is the creation of security boundaries between different business units. Notably, “business units” in this context is merely a convenient term, and it should be interpreted more liberally as any meaningful collection of computing resources (e.g., division, unit, site, region). In any case, the objective is to implement firewalls (and ideally other controls, such as intrusion prevention systems and antivirus engines) to confine malicious users and traffic to a subset of the internal computing environment — as opposed to providing wide-open access that could lead to damage to or infection of the entire environment.

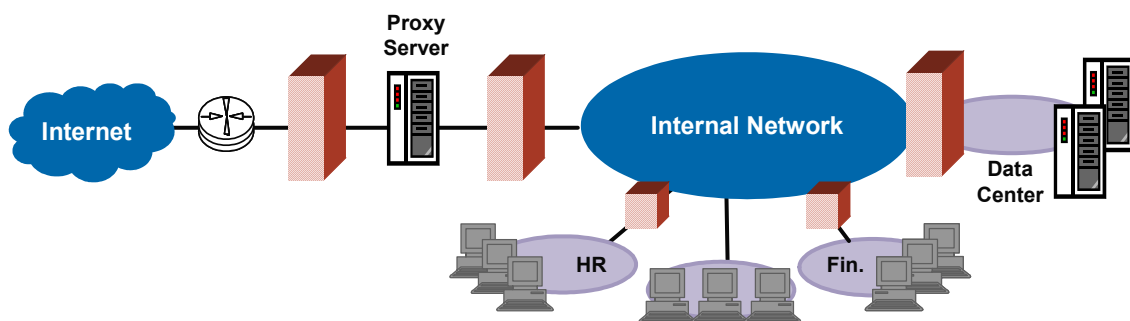
An advantage of this approach is that it inherently yields management buy-in from the local security strategy, since it enables each business unit to make its own decisions regarding how much security it requires. This approach also provides relatively good containment of threats, such as worms. However, the full realization of these benefits depends on each unit having at least a base level of in-house security skills and resources.

The Data Center DMZ

Since many organizations do not have sufficient, distributed security skills and resources, a popular alternative is to simply isolate centralized applications and resources from the general user population. Essentially, this amounts to creating a “data center DMZ,” again employing a combination of firewalls and other related controls. Although this approach may not be as effective for worm containment, it nonetheless does provide a greater degree of protection for an organization’s most critical assets. The challenge then becomes establishing a balance between having a policy that is very granular — and therefore complicated and intensive to manage — and having one that is easy to administer but too permissive. After all, there is no benefit to having a firewall if everyone is granted access to everything.

From an implementation perspective, we expect the majority of organizations will focus their investments on data center firewalling/DMZs, given the relatively favorable impact-for-cost ratio achievable with this approach. Other organizations, particularly those with a high degree of security sensitivity, will also selectively employ the practice of business-unit isolation. The combination of these approaches is shown in Figure 4, which also demonstrates how the data center firewall essentially acts as the third tier for the previously discussed two-tier, proxy-enabled Internet DMZ.

Figure 4 — Internet DMZ Supplemented With Internal Isolation



Source: META Group

The Evolution of Network Security Devices/Technology

Refining the deployment of conventional network security controls will undoubtedly yield significant benefits, but it is unrealistic to think that this alone will be sufficient to completely address current challenges relating to effectiveness and efficiency. Evolution of the specific devices and technology that comprise the various logical and physical layers of a design is also required.

In this regard, it is important to recognize that these technological advances will be applicable to and have an impact on all the aforementioned implementation options. This is the case whether the “perimeter” in question is the Internet boundary, the data center, or a business unit, or whether it involves more efficiently extending the perimeter “outward” to better account for mobile users and other external parties (discussed below).

Strengthening the “Perimeter”

In terms of improving effectiveness, we observe two main avenues of advancement for network security products. The first is the incorporation of greater application-layer control and attack-protection capabilities. Secondly, there is the simultaneous augmentation of these core services with even more extensive security functions (e.g., antivirus, virtual private networking, intrusion detection, vulnerability scanning), with the goal being to create an all-in-one network security gateway.

Taken together, we often refer to these advancements as “upward evolution,” a term that is generally intended to capture the notion of smarter, more capable solutions. However, understanding the characteristics, nuances, and limitations of these approaches is critical to maintaining effectiveness in the face of “in-progress” efforts by many of the associated vendors.

Application-Layer Awareness and Control

Historically, the predominant network security product has been the firewall, which has made access control decisions based on network-layer information. For example, traffic on communications port 25 (i.e., SMTP-based e-mail) is allowed between source (with IP address “A”) and destination (with IP address “B”), but traffic on port 80 (i.e., HTTP-based Web) is not allowed from B to A. Although this capability is useful, it is simply not sufficient in current environments, which use complex protocols and tunneling techniques.

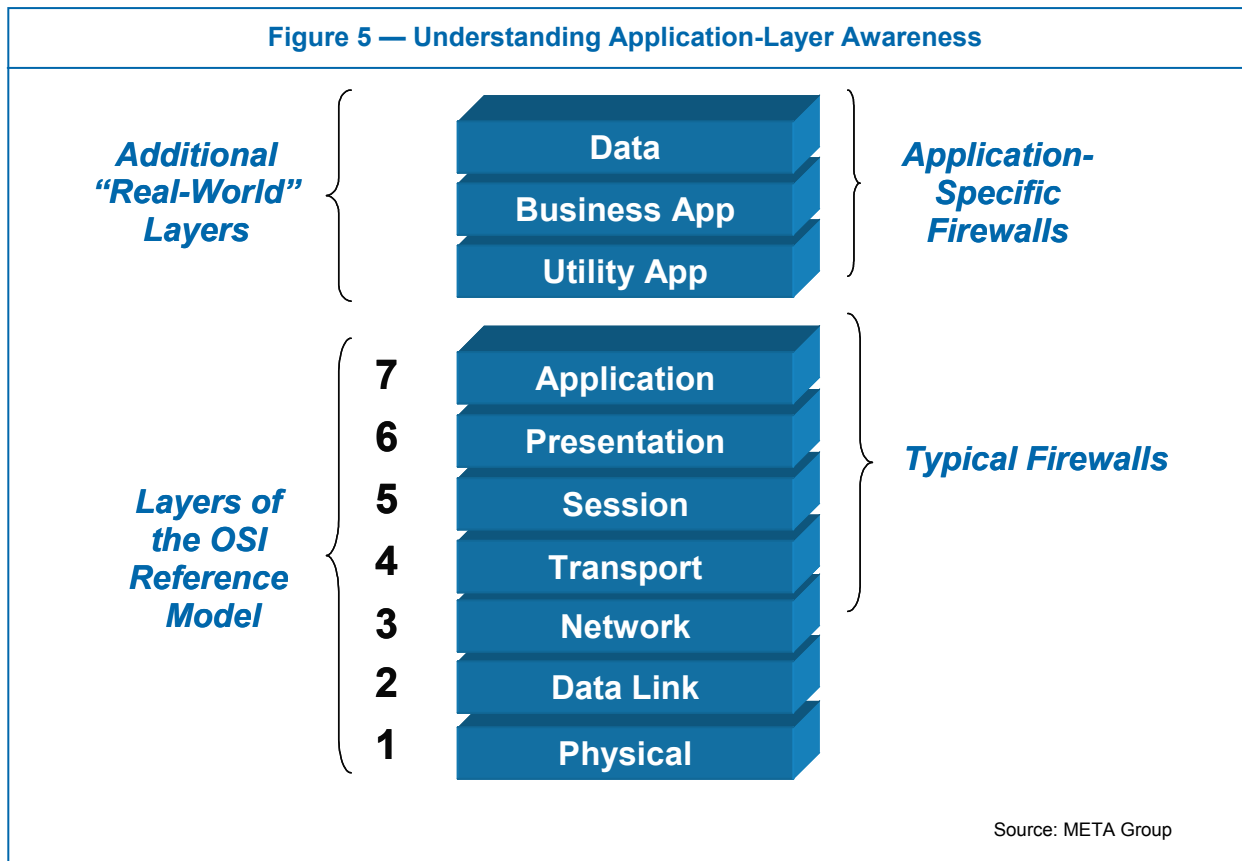
Therefore, network security vendors have appropriately begun to add more knowledge and application-layer details into their products. Fundamentally, this enables additional granularity when establishing access control policies. For example, not allowing a specific command within a given service because it is

The Evolution of Network Security: From DMZ Designs to Devices

risky, while other parts of the service are useful, safe, and therefore allowed; or not having to keep a wide range of ports “open,” since dynamic port assignments being negotiated by the communicating stations can now be tracked.

Certainly, this enhanced level of control is both good and necessary. However, it is easy to misunderstand the scope and degree of benefit. First, there is a dependency on the product vendor to add the requisite knowledge on a per-application-service basis. In other words, with hundreds of application-layer services available to choose from, it is unlikely that any one product will provide full coverage for a given organization’s communications traffic. Second, although application-layer insight is an improvement from the past, it still does not penetrate into the highest layers of the communication sessions. Specifically, it cannot operate on aspects of the application logic or individual data elements (see Figure 5). Consequently, there may still be a need in some instances to employ what we call application-specific firewalls, which have the capacity to learn and operate on these even higher layers of the communications stack. Finally, it is also important not to equate application awareness and control with the ability to prevent attacks.

Figure 5 — Understanding Application-Layer Awareness



Enhanced Attack Protection

Application awareness extends the core capability of a firewall to enforce access control on the basis of predefined permissions. However, for traffic that is allowed by policy, no other judgment is typically made as to whether it is good or harmful (e.g., contains an attack). This additional content- and context-oriented dimension of access control is commonly referred to as “attack protection,” or even “intrusion prevention.” Certainly, network security products have always had some capability in this area, but it has been very limited and predominantly based on stopping network-layer denial-of-service attacks (e.g., SYN floods).

In this regard, increased application-layer awareness is an essential ingredient to enhancing a system’s attack-protection capabilities. But, as we are apt to say, it is not sufficient. It is also necessary to employ a comprehensive set of inspection algorithms and analysis models (e.g., statistical, behavioral) to definitively establish the presence of an attack within a traffic stream. Indeed, it is the execution of these processes and their extensive consumption of system resources that pose a challenge to having combined firewall and intrusion prevention products. Although firewalling and intrusion prevention services are well aligned functionally (both being protective in nature), they are not so well aligned from the perspective of performance and system requirements.

The implication of this condition is that combined products are currently best suited for less onerous implementations, such as those within the small-and-medium enterprise market or at branch offices of larger enterprises. In these instances, lower performance demands or a subset of prevention capabilities can be deemed appropriate from a cost/benefit perspective. On the other hand, headquarters and other large sites will be better served by separate, dedicated intrusion prevention systems that incorporate a full range of prevention mechanisms. That said, by 2005, we expect leading network security vendors to deliver high-performance platforms capable of both full-featured firewalling and intrusion prevention, which should be suitable for implementation in any scenario.

To summarize this aspect of technological evolution, increased application awareness is a foundational capability that must be demanded of all network-based security products. Furthermore, organizations must take care not to be enamored or misled by catchy marketing terms, such as “deep packet inspection,” “application intelligence,” “application defenses,” and “total stream inspection.” Regardless of the terminology being used, the objective should be to obtain *both* greater degrees of permissions-based control *and* enhanced attack protection (see Figure 6).

Figure 6 — What About Intrusion Detection Systems?

The intent is that attack protection capabilities in the form of intrusion *prevention* systems (IPSs) pick up where IDSs (intrusion *detection* systems) historically have left off. Specifically, whereas IDSs identify suspicious activity or probable attack scenarios, the goal of an IPS is to definitively establish the presence of an attack and then stop it. For the most part, this distinction hinges on the accuracy and real-time nature of the underlying inspection, detection, and analysis engines.

We envision an important role for both of these security services:

- IDSs will serve in an operational audit capacity, verifying that protective controls are operating as intended and continuing to provide alerts and information pertaining to suspicious events.
- IPSs ultimately will be combined with firewalls to form the basis of a new breed of protection-oriented security gateways.

Source: META Group

Multiservice Security Gateways

Similar to combining firewalling and intrusion prevention, the goal of a multiservice security gateway (MSG) is to enhance effectiveness by bringing more security services together into a single physical device. It is important to note that this approach also has tremendous potential for improving efficiency in that it may reduce the need for numerous standalone products and their supporting infrastructure (e.g., switch ports, databases, management servers). However, also comparable to the firewall/IPS scenario is the reality that this concept is not without some potential issues.

We recommend that organizations evaluating MSGs take into account the following considerations and criteria. An ideal solution should:

- **Provide a reasonable, expected cost reduction (e.g., at least 30%) over the use of a collection of point products:** This is necessary to ensure that an *actual* cost savings is realized, to account for the costs/disruption associated with making a change, and to account for any potential shortcomings or compromises associated with the solution being a multiservice security gateway (see following bullet items). Furthermore, it is important not to overestimate expected savings in the area of ongoing, operational management. Some multiservice solutions consolidate point functions, but require them to continue to be managed separately.
- **Provide a set of services that are either best of breed or a best fit for the organization:** It is unlikely that compromising on quality to achieve initial cost savings will ever prove to be a worthwhile approach.

- **Provide a set of services that is functionally well aligned or that is at least sensible for supporting a specific use-case scenario:** This relates to understanding the degree of practical usefulness of a multiservice security gateway.
- **Provide a set of services that the organization actually needs:** Excessive overlap with previous investments can be an issue. Therefore, MSGs with selectable service sets (versus fixed) should have an advantage.
- **Be secure:** Having additional services is generally positive in terms of security effectiveness, but not if it presents an opportunity to short-circuit a defense-in-depth strategy (e.g., as a result of design flaws or configuration errors that allow services to be bypassed under specific conditions). Admittedly, this scenario is only remotely possible, and therefore should be treated as a secondary concern. However, ultimately, security effectiveness is something that is demonstrable only by a product standing the test of time. Therefore, prudent organizations are advised to wait at least a year before deploying newly released multiservice security gateways in highly sensitive use cases.
- **Provide adequate performance:** It is important to establish that performance objectives can be met when all necessary services are operating and processing real-world traffic with real-world configuration settings. (Large numbers of large-packet UDP sessions all doing the same thing do not provide the basis for a truly indicative performance test.)
- **Be manageable:** With multiple services running on one device, it is essential that the associated management application be able to scale (e.g., by supporting delegated administration) and be able to avoid violation of the least-privileges principle (e.g., by supporting separation of duties via role-based administration).
- **Be integrated:** This criterion is treated last because it directly impacts each of the three previous items: security, performance, and manageability. MSGs that are not truly integrated and that are instead “virtually separate devices” will not reap the benefits of shared-packet processing, shared event information, and shared management capabilities.

At the risk of sounding like a broken record, drawing even further parallels to the combined firewall/IPS scenario is warranted. Specifically, due to the above list of considerations, we expect that multiservice security gateways will initially have the greatest applicability for small and medium enterprises and branch offices. In addition, they will also gain traction in specific, targeted use cases, where the combination of services and the available performance are just right, such as a security backstop (e.g., firewall, antivirus, IPS) to a remote access server/concentrator or to a dedicated partner connection.

Yet as these solutions prove themselves, particularly in terms of security effectiveness, performance, manageability, and achievable cost savings, we expect adoption on a more wide-scale basis, with “well-heeled” products available by 2006. In terms of the impact on the overall design or architecture vis-à-vis multi-tier DMZs, we expect these solutions will be used predominantly to combine all the security services within a given tier, but generally not to collapse multiple tiers into a single device.

Extending the “Perimeter”: Outward Evolution

From a high-level perspective, the following section is meant to convey the need to better account for the extended enterprise in the form of mobile employees and third-party users with substantial degrees of access or interconnectivity. Treatment of this topic arguably could have been placed in the first section of this paper (“The Evolution of Network Security Design”), which focuses primarily on enhancing security by improving the location and arrangement of various controls. If our intent had been to discuss endpoint security mechanisms, such as personal firewalls and antivirus agents, then that would have been the case. However, the technology we intend to focus on lies at the intersection of numerous themes, and thus it is appropriate to cover this topic here. Specifically, SSL VPNs enable the computing environment to be securely extended while delivering benefits over previous approaches in terms of all three of our aforementioned critical objectives: efficiency, flexibility, and security effectiveness.

The SSL VPN Revolution

In the past, the predominant mechanism for achieving secure remote access has been IPSec virtual private networking technology. Yet IPSec has been hindered by the burden of having to deploy, manage, and maintain a software component on each node that needs to communicate. It has also been impacted by the inability to effectively provide access that is more granular (i.e., less “wide open”) than to an entire network. As a result, most organizations have constrained their usage of IPSec remote access solutions to a relatively small portion of their user population.

In contrast, SSL VPNs take advantage of ubiquitous browsers and dynamically downloaded modules to achieve the client end of an encrypted session. This introduces a great deal of flexibility, since it relieves the limitation of users being restricted to only those computers with pre-installed client software. User-owned or partner-owned computers and even Internet kiosks can now be leveraged to provide secure communications.

Further advantages, capabilities, and characteristics of a leading SSL VPN solution include:

- **A multipart architecture:** The focal point of this is an SSL-enabled gateway that is typically located within an Internet DMZ. Other elements of the architecture include the client (i.e., a browser), a management application that addresses both policy configuration and system monitoring, and a separate policy store (optional).
- **The ability to provide access to a comprehensive set of applications:** Typically, three connection modes are used to accomplish this. The first relies solely on a browser as the client and provides access to both Web applications and a subset of other applications for which the vendor has built HTTP translation capabilities directly into its gateway (e.g., standards-based e-mail, file services). The second mode involves dynamically downloading a “plug-in” module to the client browser, and it supports access to well-behaved client/server applications (i.e., those with well-known, static port definitions). Finally, the third mode similarly involves a downloadable module and provides full network access that is essentially equivalent to that delivered with IPSec.
- **A wide variety of security-related features:** These include the following:
 - At least with the first two connection modes, access is limited to a specific application session, versus an IP address, subnet, or entire network.
 - At least for the first connection mode, even more granular control should be possible. We refer to this as “authorization,” and it involves being able to specify which subapplications or individual functions are available to a user.
 - The gateway itself, which often incorporates some degree of proxying and protocol translation services, can inherently support a high degree of additional filtering (e.g., for viruses or for protocol conformance). However, we note that this is an area that is generally underdeveloped in most currently available products.
 - The degree of access can be established dynamically based on a wide variety of attributes (e.g., state and strength of authentication, security audit of the client machine, type of client platform).
 - Highly granular log information should be obtainable.
 - In many instances, it should be possible to institute cache cleaning on the client station when a session is completed. Increasingly, this capability is also being expanded to enable the creation of a secure, virtual workspace on the client machine.
 - Single-sign on (SSO) and/or integration with leading Web-SSO products is becoming common.

Given this range of capabilities, it is also appropriate to offer two additional recommendations. First, the high degree of flexibility and broad set of

configuration options make it an absolute necessity to have a very robust and efficient management interface/system. Second, given the degree of proxying, encrypting, and application-level processing that will typically be conducted, special attention should be paid to the availability of performance-enhancing characteristics (e.g., special-purpose hardware, caching, clustering capabilities).

From an implementation perspective, we expect adoption of SSL VPNs to continue to accelerate. By 2006, it will become the dominant approach for achieving secure remote access, with greater than 70% of all users employing it as the method of choice. Furthermore, its potential for use in intranet environments should not be overlooked. Indeed, a ubiquitous implementation of SSL VPN on an internal network could conceivably relieve the need to deploy business-unit barriers — though use of data center DMZs would likely remain a necessity.

Bottom Line

Increasingly sophisticated cyberattacks, business and economic requirements, and even regulatory considerations are driving changes to both network security devices and their physical and logical arrangement. To ensure effectiveness, efficiency, and flexibility, solutions must evolve inward, upward, and outward. Organizations must acknowledge and embrace these changes — or accept the minefield of negative outcomes that accompanies mismanaging the risks associated with operating a modern-day computing environment.

Mark Bouchard is a senior program director with Security & Risk Strategies, a META Group advisory service. Brian Golumbeck is a director with META Group Consulting's Operations, Infrastructure, & Security Practice. For additional information on this topic or other META Group offerings, contact info@metagroup.com.



About META Group

Return On IntelligenceSM

META Group is a leading provider of information technology research, advisory services, and strategic consulting. Delivering objective and actionable guidance, META Group's experienced analysts and consultants are trusted advisors to IT and business executives around the world. Our unique collaborative models and dedicated customer service help clients be more efficient, effective, and timely in their use of IT to achieve their business goals. Visit metagroup.com for more details on our high-value approach.

