NOKIA
CONNECTING PEOPLE

The Next Network Challenge: Building A LAN Infrastructure Designed for Outsourced Data Services

A White Paper from Nokia Internet Communications

January 18, 2000

Not too long ago, a disruption of a company's Internet or WAN connectivity meant delays in receiving email, and some minor inconvenience to employees wishing to surf the Web. Now, with email, groupware, ERP, and even basic productivity applications provided over the Internet, a break in Internet service can be as severe as – or even indistinguishable from –a crash in the local area network.

That's because an enterprise network does not stand alone. Conventionally, the border between the enterprise LAN and the Internet is clear-cut: the WAN router. But thanks to the rise of the Internet as a resource for business, and the increased popularity of outsourcing network services, the boundary is increasingly blurred. No longer can companies worry just about the integrity of their own local area networks and of the services maintained on it.

In this White Paper, Nokia will describe the evolution of the enterprise network from a discrete entity to one tightly coupled with the Internet, and the implication of that evolution to an IT director's greatest concern: application availability. We will then present our vision for a seamless, scalable architecture which ties together the enterprise LAN, the Internet Service Provider, and application services providers' networks. The goal: to increase system and application availability, as well as provide a fault-tolerant infrastructure which can transparently recover from the inevitable temporary service disruptions, without impacting the end user.

The WAN, from Batch to Real-Time Business Critical

In the dark, distant past, the WAN was a convenience. Five to ten years ago, many enterprise LANs were of a hub-and-spoke design. At the center of the network sat centralized resources, such as a WAN router, centralized authentication servers, database servers, file/print servers, and even applications servers offering email, groupware, or other applications. Connections to legacy systems, such as minicomputers or mainframes, were made at that central point.

Most traffic on the network passed from end nodes, or client workstations, located around the perimeter of the network, in toward the center. In a large facility, departmental servers handled local traffic, while centralized resources were used to back up those servers or provide enterprise-wide services.

In the case of a large business with multiple locations, each remote location had its own hub-and-spoke network, consisting of core server and with X.25, fractional T-1, or Frame Relay WAN connections, with the perimeter consisting of nodes requiring access to those resources. The WAN connections were often used to provide low-bandwidth synchronization of remote locations' servers back to the headquarters data center's central core, or to exchange email with an Internet gateway. Individual users weren't even aware of those WAN links; their network resources were entirely local.

In order to maintain high availability in those environments, network managers had three priorities:

* First, to ensure the stability of the individual servers, meaning the hardware, software, and data, in particular to ensure that data would not be corrupted.

*Second, to provide fault-tolerant and redundant links between end users and the central resources.

*Third, to guarantee the availability and predictability of the WAN links between remote offices and the headquarters data center.

The WAN connection, however, was not essential for real-time operations and maximizing end-user productivity. The WAN was used for batch-oriented data transfers, providing movement of email or database changes between locations. Short-term WAN service disruptions – between five minutes and perhaps a few hours – might not even be noticed by employees, lessen their productivity, and affect the business' bottom line.

Yesterday, the WAN was useful, but not critical. Over the past two or three years, in many organizations, there was a distinct change in the nature, quantity, and priority of traffic traversing the wide-area network, either over a private WAN link to another facility, or to the public Internet. The most obvious change in the traffic pattern is the greatly increased email traffic exchange between the enterprise's email servers and an ISP's SMTP gateway, as well as HTTP both from employees' browsers outward to the Web, and from customers and business partners coming in to the enterprise's publicly accessible Web servers.

Other traffic, traversing private links such as frame relay and T-1, would tie remote offices to the headquarters' databases and groupware in real time, but the expectation was that these WAN links would not be real-time business essential. Applications such as Lotus Notes provided facilities for replication of data across the wide area. A short-term loss of WAN access might delay the delivery of email or of database replications, but again, those would not be business critical failures. Employees' short-term lack of access to the World Wide Web, back in 1996 or even early 1998, was more of an inconvenience than a crisis.

A bigger problem was that of non-employees – potential customers – not being able to access internally hosted public Web servers, but before the advent of direct-sales e-commerce, the occasional "404 Not Found" was an expected inconvenience for a Web surfer, and rarely if ever meant the loss of revenue.

Today, the WAN is pervasive for individual users. Internet traffic represents the lifeblood of many businesses, and Web sites represent a key vehicle for sales and marketing. If a customer clicks on a banner ad and sees "404 Not Found," that's not only a lost opportunity for sales, but also a waste of marketing resources. For companies that host their own Web sites internally, the integrity of the Internet connection is vital.

Over the past year or two, however, businesses have come to rely upon real-time links, over the Internet or over private IP-based networks, for provisioning basic business functions. An easy starting point for many enterprises wishing to focus on business-critical tasks is to outsource their email functions to their local Internet Service Provider, or in some cases, to specialized Application Service Provider.

End users' accessed remote servers for all of their email needs, whether by using a Web-based interface, or familiar POP3/IMAP4 email clients such as Microsoft's Outlook, Netscape Communicator, or Qualcomm's Eudora Pro. Without a reliable WAN link, users lost more than speedy delivery of extra-company email – they lost access to their entire email infrastructure, including message archives and intra-company messages.

Tomorrow, the WAN link will be key to many business functions. The trend toward application outsourcing began in 1998, and gained momentum in 1999. Moving beyond mere email and public Web sites, many enterprises found that although they needed a standard application, such as Lotus Notes, Microsoft Exchange, or Oracle 8, they didn't want or need to physically maintain the servers.

With Web-based management interfaces, high-bandwidth WAN links over private IP networks or even using virtual private networks (VPNs) over the Internet they could have the benefits of owning the applications, but without the need to own the hardware, or maintain a crew of increasingly expensive and hard-to-retain certified application experts. Even Enterprise Resource Planning (ERP) applications, such as SAP's R/3, became available over WAN from service providers. A new breed of

Web-based business-to-business e-commerce tools, based on XML technology, will also be outsourced over the Internet. Another hot trend, expected to catch on in mid-to-late 2000, is productivity application rental, being pioneered by companies such as Sun Microsystems with its purchase of the Star Office application suite.

The nature of these business-critical application requires non-stop access to them, not only by employees, but frequently by customers, supplier, and partners. As such, the service provider's data center becomes an increasingly attractive location for such servers, applications, and data to reside, due to its closer and high-bandwidth ties to the Internet backbone.

The Nature of the Outsourced Applications

Two interesting factors differentiate public Internet access – checking email, surfing the Web, buying books at Amazon.com, checking stocks at E*Trade – from the business-critical outsourced WAN applications described above.

The first is that, in the case of business-critical outsourced applications, high availability cannot be sacrificed, and network reliability is as important as the quality of the server hardware and software itself.

The second is that authentication and encryption needs are paramount, and the mechanisms for creating and maintaining authenticated/encrypted links over the WAN are much more sophisticated than required for consumer e-commerce.

Consider a typical scenario for a modern enterprise in late 2000 or early 2001. A company with 300 employees in three locations: 200 in a headquarters office, and 50 each in two field offices. Half of all employees are using an ERP solution hosted at an application service provider, or ASP. Half are also using an application suite rented from an ASP. Half of the field-office employees are connected via virtual private network back to the headquarters office. The company also operates a local e-business solution, which is in use by 20 business partners, as well as 10 employees from each field location.

From the headquarters site, there will therefore be, during normal business hours, 240 highly encrypted and authenticated sessions. That consists of 200 outgoing sessions: 100 from employees to their ASP's outsourced ERP solution, 100 from employees to the ASP's outsourced application suite. 40 encrypted/authenticated sessions will be incoming, 20 from partners, and 10 each from field offices.

Each field office will similarly have 60 encrypted and authenticated sessions in progress: 25 to service provider's ERP application, 25 to the basic productivity application, and 10 back to the headquarters' e-business solution.

The Nature And Fragility of the Encrypted/Authenticated Sessions

Because our sample company rightfully takes data security seriously, as do its partners and ASPs, a sophisticated combination of X.509-compliant digital certificates and the IPSec secure tunneling protocol are used to authenticate each session. Once sessions are authenticated, encryption is performed using one of several high-end protocols, such as 156-bit TripleDES.

For access to its mainframe e-business application, the enterprise maintains its own in-house certificate server, used by both field employees and the company's business partners. Access to the outsourced productivity and ERP applications is validated against the ASP's own certificate server.

In either case it takes a small, but measurable amount of time to authenticate users via their digital certificates, and then create the IPSec encrypted session tunnels which provide user access to the various applications. This may be in the range of five to 15 seconds if all parties are using hardware-accelerated encryption technology; maybe three or four times that if all security operations take place in software. This slight delay does not typically represent an inconvenience for users, as they generally remain logged into their applications for hours at a time.

A challenge with secure links are their relative fragility – relative to the local-area network itself. A service breakdown can occur at any point in the data chain, whether the secure connections traverse the Internet, or remain within the boundaries of the enterprise LAN. The addition of secure authentication and encryption doesn't make the link more fragile, but it *can* exacerbate the effects of a service outage, because if the one of the devices in the secure link fails, the entire secure connection fails, and must be reestablished from scratch.

Assume that the company used in our example has two secure network pathways between individual users and network resources like the ASPs and its inhouse electronic commerce servers. Each one of those pathways uses some type of security devices for authenticating or encrypting secured connections to the business's mainframes. That device might be a software- or hardware-based VPN server attached to the enterprise's switches or routers, an IPSec gateways, or some other critical gear with the necessary authentication and encryption capability. Thanks to load-balancing hardware, each of those links is used to handle one-half of the company's 240 secure sessions, as described earlier, for a load of 120 sessions per security devices.

Now, assume that one of those links failed, for any reason ranging from a power failure to a fault in the network. It makes no difference where that failure happened – the result will be the same, as that link's 120 sessions will be broken. Half of all users will lose their work in progress. They will need to make a new contact via some other path in the company's networking infrastructure, authenticate themselves via X.509 digital certificates, and establish IPSec-compliant encrypted packet tunnels, and then reconstruct their work-in-progress.

Depending on the performance of the data links, the certificate servers, and the authentication/encryption hardware and software, this could produce a considerable delay before the employees and partners can begin useful work.

At this point, of course, one authentication/encryption system is now handling the company's full load of 240 sessions – even after the original link is returned to service. Over time, as users log out and log back in, a balanced load will return. But for the present, the company's users not only will experienced reduced performance, but the system will be less fault-tolerant, and less able to withstand and gracefully recover from any possible second service outage.

Now imagine this same scenario for a larger enterprise, with tens of thousands of employees in dozens of locations, all connecting to each other, to partners and suppliers, and ASPs over the Internet and private IP networks. The consequences of even short-term failure of secure connections would be significant.

Eliminating Failure With Clusters

In this outsourced and network-centric world, enterprises cannot afford the loss of productivity associated with inevitable service outages or hardware failures, as well as an inability to immediately regain full fault tolerance once failed equipment is returned to service.

Of course, this is not a problem unique to authenticated and encrypted services. Enterprises have implemented solutions to similar problem encountered with servers, with local-area network gear, and with storage arrays. The solution: *Clusters.*

Generally speaking, a cluster is a collection of interconnected and functionally identical computing devices. To the outside world, a cluster acts like a single computing device with a single IP address. One member of the cluster acts as the *cluster master*, and takes charged of distributing the workload between all operational computing devices in the cluster, called the member *nodes.*

Each node in a cluster not only knows about its own current secure sessions, but maintains detailed information about the current state of the workload of each of the other nodes in the cluster. Thus, If one node malfunctions, the cluster master reapportions its workload to the other nodes, which can continue the failed node's transactions instantly. If the cluster master fails, another node assumes that function in a matter of milliseconds. The benefits of clustering is that the workload is balanced between a number of devices, and that as long as at least one node in the cluster continues to function, work can continue – there is no single point of failure.

A cluster's fully redundant architecture also is scalable. If the work load increases, it's a relatively simple matter to add additional nodes to the cluster. As soon as a new node goes online, the cluster master puts it to work.

Many IT professionals are familiar with the concept of server clustering. Server clusters are common at large Web-server farms, where incoming requests for Web pages are direct to the cluster, not to individual servers, and the cluster automatically apportions the workload between all active members of the cluster.

Clustering technology has applications far beyond server clusters. They can also be used to provide fully redundant and scalable processing of network session authentication and encryption. Consider the company headquarters described above, with the two authentication/encryption systems. Imaging if those security devices were clustered together, invisibly integrated into a single logical unit which appeared to be a single fault-tolerant security service device.

Under normal operations, each security device – which is now a cluster node – would handle one-third of the company's 240 secured sessions, or 80 sessions. In the event of a failure of one node, the cluster master would allocate the 80 open sessions between the two existing nodes, giving each a load of 120 sessions. A key factor is that the failed node's sessions would not be broken – because each node maintains a copy of the state information for all of nodes in the cluster, they can simply take over the existing sessions without losing a beat. And when the third node is repaired and placed back in service, the cluster master can immediately reallocate the work – and each node would be back to its original 80-session workload.

At present, there is no industry-wide standard for clustering authenticated and encrypted sessions. There aren't even standards for dynamically load balancing or for offering automatic failover of those sessions without loss of their current state. For the present, there are only proprietary solutions, such as those offered by the Nokia CryptoCluster IP clustering technology and product family.

As enterprises investigate and deploy outsourced solutions to business-critical tasks, we urge them to consider the productivity issues inherent in increased reliance upon authenticated and encryption network links. We also urge other manufacturers of remote access, virtual private networking, wide-area routing, and related technologies to consider the issues inherent in reducing customer pain, and to work with us to promote industry-wide awareness and standardization of secure-session clustering, and automatic failover without loss of session state.

At Nokia Internet Communications, our customers are depending on it.