NO WARRANTIES OF ANY NATURE ARE EXTENDED BY THE DOCUMENT. Any product and related material disclosed herein are only furnished pursuant and subject to the terms and conditions of a duly executed license or agreement to purchase or lease equipment. The only warranties made by Unisys, if any, with respect to the products described in this document are set forth in such license or agreement. Unisys cannot accept any financial or other responsibility that may be the result of your use of the information in this document or software material, including direct, indirect, special or consequential damages.

The information contained herein is subject to change without notice. Revisions may be issued to advise of such changes and/or additions.

Unisys is a registered trademark of Unisys Corporation
Contents

Introduction.............................................................................................................................................. 4
A Typical Systems Architecture to Support E-Business ........................................................................... 6

  Global Server Load Balancing ............................................................................................................. 7
  Firewalls and Intrusion Detection Systems ........................................................................................... 7
  Firewall Load Balancing ....................................................................................................................... 7
  Caching .................................................................................................................................................. 7
  Reverse Proxy Servers ............................................................................................................................. 8
  The De-Militarised Zone (DMZ) ............................................................................................................. 8
  The Web Server Farm and Server Load Balancing Technologies ........................................................... 9
Summary.................................................................................................................................................. 12
Introduction

Systems Infrastructures that support e-Business combine a number of technology solutions into a complex environment necessary to provide a secure, highly available, highly resilient, 24x7 service.

Technologies include wide area networking, global server load balancing, data caching, firewalls, intrusion detection systems, reverse proxy servers, server load balancing, de-militarised zones, SSL accelerators, web server farms and the management processes and procedures that are necessary for any systems solution. These technologies move at an alarming rate, and one of the issues facing a business implementing, or even a business with an existing e-business infrastructure is keeping abreast of the technological change.

Unisys do this through their unique SystemFlow Methodology. This is a proven methodology for the design and implementation of systems infrastructures; the main difference between SystemFlow and other methodologies being that it has an infrastructure design and delivery focus, rather than an application design and delivery focus.

The SystemFlow Methodology

SystemFlow also underpins Unisys Technology Consulting Services ISO9001 accreditation. It is the means by which Unisys develop, capture and distribute best practice in the design, implementation and management of systems infrastructures to its consultants throughout the world.

Since its release four years ago, SystemFlow has been expanded to include best practice in the form of Resource Kits. These focus on the implementation of technologies that are key to the successful delivery of projects to meet specific client requirements, for example Internet and Intranet technologies. The best practice documents are updated regularly to include details of the very latest products and implementation techniques from the field.
They are available to our consultants enabling them to advise and guide our clients in the development and implementation of systems infrastructures that are able to meet their business requirements.

This white paper provides a high-level introduction to the technologies necessary to build an e-business infrastructure; these are dealt with in detail in our SystemFlow e-Business Resource Kit.

The SystemFlow e-Business Resource Kit
A Typical Systems Architecture to Support E-Business

The diagram below shows a simplified ‘typical’ e-business systems infrastructure. Even simplified it looks fairly complex.

A Simplified Diagram of an E-Business Infrastructure

Technologies included in the above environment are very specific to Internet and Intranet environments, looking at each technology in turn:
Global Server Load Balancing

This provides for the balancing of client sessions across two or more physically distributed sites for purposes of resilience and disaster recovery purposes. Due to the 24x7 nature of the Internet, it is becoming increasingly common for providers of an Internet service to operate two or more sites in an active-active configuration.

There are many ways to balance client sessions across this type of environment, from a simple DNS round-robin configuration, where two or more IP addresses are configured for each URL on the DNS server and clients are connected to each in turn as the DNS request is resolved, to DNS re-direction based and BGP based intelligent load balancing supported by load-balancing switch manufacturers such as Cisco and Nortel.

Firewalls and Intrusion Detection Systems

Firewalls impose an access policy between two separate networks, such as the Internet and internal network of an organisation. They inspect every packet of information to determine whether the packet should be allowed to pass freely between the networks, or simply ignored. The information to determine this is stored within the firewall rule-base.

The rule-base itself should be kept small; in general to no more than 15 to 20 rules. There are two reasons for this: the first is that the larger the rule-base, the more processing is required within the firewall, and the more of a bottleneck the firewall becomes; the second, and the one that is the biggest risk to the business, is that the larger the rule-base, the greater the chance of having conflicts in the rule-base that inadvertently allow unwanted ‘hackers’ into the internal network.

It is also worth considering that the same rule-base associated with a firewall, can usually be applied to a router, and that the only thing that differentiates the two is that a firewall has a log. This log is critical, because a firewall is only as effective as the process to monitor and act on events in the log. If you do not monitor the log regularly, you severely diminish the effectiveness of the firewall.

The Intrusion Detection System (IDS) works in ‘partnership’ with the firewall. Probes are inserted at various points on the internal network; the De-Militarised Zone (DMZ), behind firewalls, proxy servers and on the internal network itself. The IDS monitors traffic on the network and looks for specific profiles that indicate that malicious activity is likely and can alert and/or automatically respond to such an event.

IDS systems can interact with the firewall to close a specific session or port to minimise the effect of any malicious activity. IDS systems have standard profiles configured. However, the real power of an IDS can only be realised by fine tuning these profiles with information based on your particular environment, and the profiles of the types of people who are deemed the greatest threat to the service or network.

Firewall Load Balancing

Redundancy within an e-business systems infrastructure is necessary in order to provide the levels of availability required for a 24x7 service. Redundancy of the firewall can be addressed in several ways; using products like Stonebeat to provide an active-standby configuration, using products with VRRP support, or using load balancing technology.

The benefit of using load balancing technology is that the firewalls are able to operate in an active-active configuration. Furthermore, if additional bandwidth is required, you can add a further firewall, increasing performance whilst managing the physically separate firewalls as a single logical unit with a single rule-base.

Load balancing technology also provides a layer of isolation between the Internet and firewall. From the Internet side of the load balancing switch, the only visible IP address is that of the virtual IP address used for load balancing; the addresses of the firewalls themselves are invisible. Furthermore, the load balancing switch can be configured to only allow traffic from the Internet through on defined ports, usually port 80 (HTTP) and 443 (HTTPS).

Wherever there is a load balancing switch in the infrastructure, you effectively add an additional layer of security.

Caching

As the functionality and demand for your Internet service grows, there is a greater demand on the infrastructure components. To minimise this demand, and to reduce the impact on performance of bottlenecks such as firewalls and reverse proxy servers, it is usual to deploy cache servers.
Cache servers serve static content that can make up over 80% of the content served from a typical e-commerce site.

Cache servers can be deployed in two locations within the infrastructure; the DMZ, and in front of the firewall, usually served by a load balancing switch. In the early days of server load balancing technology, switches were often termed ‘layer-4 switches’, because they worked at layer 4 of the OSI stack. Current load balancing technologies work at layers 4 through 7, that is, all the way from the transport to the application layers.

This enables them to understand what content the client is requesting from the web servers, and, if that content is recognised as static content, for example, gifs or jpegs, it can redirect the request to a cache server to satisfy that request, rather than through the firewall to the reverse proxy and web servers. This can dramatically improve the performance of the service provided.

Within the DMZ, content is often cached on the reverse proxy servers.

Reverse Proxy Servers

Reverse proxy servers are effectively application firewalls. Normal firewalls allow network traffic through based on the originating and destination IP address, and the port. The rule-base of the firewall defines what external IP addresses are allowed to talk to what internal IP addresses over defined ports, such as HTTP (port 80) and HTTPS (port 443).

Reverse proxy servers allow communication between the client and the web service based on what URL the client is requesting. Furthermore, the client is not allowed to talk to the web server directly, the reverse proxy will retrieve the information or pages requested and return them to the client on behalf of the web server. There is no direct client to web server session or communication.

The reverse proxy server is also able to redirect requests to web servers, on different ports and IP addresses from those requested, offering a form of Network Address Translation. The effect of this is that requests for multiple web sites on port 80 and port 443 of the reverse proxy servers, can be redirected to differing web servers and services on differing IP addresses and ports on the internal network.

This allows the firewall between the reverse proxy servers and the internal network to be closed to any traffic arriving at the firewall on ports 80 or 443. Preventing any possibility of tunnelling through the DMZ on these ports, which are open from the Internet.

Another feature of the reverse proxy server is the ability to cache static data, in exactly the same way as cache server technology deployed in front of the firewall. This is useful when SSL is used to provide a service as encrypted pages cannot be served from a cache outside of the firewall. SSL can be decrypted on one side of the reverse proxy allowing static content to be served from the proxy cache. Onward traffic can then be re-encrypted if necessary, although it is common for the internal session between the proxy server and web server to remain unencrypted.

The De-Militarised Zone (DMZ)

The DMZ is a protected network located between the Internet and internal (private) network. The DMZ forms a buffer zone for all traffic entering the internal network from the Internet, such as HTTP and HTTPS traffic.

The DMZ comprises a separate sub-net with physically separate firewalls (of differing types to further improve security) on either side of the network. The firewall rules do not allow external requests to pass directly to the internal network; they must pass though servers located in the DMZ. This provides a layer of security for the internal network.

Furthermore, traffic specific to the Internet (e.g. external DNS) is prevented from entering the internal network as Internet DNS services can be located within the DMZ if not hosted externally.

There are several types of servers that are commonly placed in the DMZ:

- Internet DNS servers.
- Mail and HTTP content-scanning servers (for protection against viruses or malicious ActiveX controls, and for scanning of keywords, addition of disclaimers, etc).
- Reverse-proxy servers.
- WAP gateways.
Typically, when implementing a DMZ, the logical firewall between the Internet and DMZ is configured to use NAT (Network Address Translation) to translate the advertised external IP address(es) for the site to internal addresses on the DMZ. This shields the internal addresses within the DMZ from the Internet.

When doing this, it is common for the DMZ addresses to be in the range from the private address ranges, as defined by IANA (Internet Assigned Numbers Authority). The private IP address ranges are defined within RFC1918 published by the Internet Engineering Task Force (IETF). These IP addresses will never appear on the public Internet and can be used for internal private networks without fear of them ever appearing as a site. Further information on RFC1918 can be found at [http://www.internic.net](http://www.internic.net).

The reason for doing this, is that as these addresses will never appear on the Internet, they cannot be routed over the Internet, and it makes ‘spoofing’ (where an external user attempts to identify themselves as an internal user by mimicking an internal IP address) more difficult (it is worth noting that firewalls also have anti-spoofing capabilities, but implementing the DMZ within the internal private network address space makes it even harder for hackers).

If we consider our infrastructure described so far, we have many layers of protection between the Internet and internal network, and the ‘old’ concept of the DMZ differs greatly from that of a modern system infrastructure:

- The global server load balancing switch prevents external users from directly accessing the firewall IP address(es), and will only allow traffic through on ports for which load balancing has been configured, usually ports 80 and 443.

- The firewall is configured to translate external IP addresses to the addresses within the DMZ. These addresses are within the address range defined by IANA as private address space, and are not directly routable from the Internet itself.

- Internal load balancing switches within the DMZ balance requests across the reverse proxy server farm. Once again, only the virtual IP address is visible from the firewall, the physical addresses of the proxy servers are not directly addressable, and only traffic over ports 80 and 443 are allowed to pass.

- The reverse proxy servers prevent client requests from directly accessing the web servers by retrieving the data from the web servers on behalf of the client. There is no direct session between the client and the web server; a session exists between the client and reverse proxy server, and a separate session exists between the reverse proxy server and the web server.

- The reverse proxy server also translates the request from the client for a particular URL and page to a physical web server IP address and virtual port. The virtual port over which the web service is served to the reverse proxy server is not port 80 (HTTP), but commonly ports 8080, 8081, 8082, etc., allowing multiple web sites/services to be served from a single web server farm over a single Internet IP address.

- The firewall between the DMZ and the internal network is configured to reject traffic trying to enter on ports 80 and 443 (the ports open to the Internet), and the firewalls themselves are shielded from the reverse proxy servers by another load balancing switch.

These multiple layers of security at the network layer make unauthorised access at the network layer to the internal network from the Internet at the worst very difficult, and at the best, pretty much impossible.

### The Web Server Farm and Server Load Balancing Technologies

Once into the internal network, the web server and application server farm can be deployed. Until fairly recently, it was common to site the web server farm within the DMZ. However, there were several issues with this:

- Having the web servers behind a firewall makes implementing management functions such as content management more difficult. Opening the firewall between the DMZ and internal network for this purpose weakens security.

- Remote access to the web servers for management purposes was also difficult without further weakening security; support staff often had to go to the physical server to manage it.

- Implementation of Microsoft’s DCOM protocol was made more difficult as this is not naturally confined to a small number of ports; to implement the protocol you had to restrict the ports it used, or tunnel it through HTTP (the same is also true when implementing other protocols, such as Oracle’s NET8 protocol).
With the implementation of reverse proxy servers within the DMZ, and the isolation from the client provided by the reverse proxy servers, it became feasible to relocate the web servers within the internal network. This allows management functionality, such as content management, release management and maintenance to be implemented more easily.

Web server farms are considered the most appropriate method of implementing a large web service capability (although server consolidation in this area will become more feasible over time). A technology that is key to the success of a web server farm is server load balancing technology. Load balancing technology has already been discussed within the infrastructure to provide global server load balancing, firewall load balancing and proxy server load balancing, but its original purpose was to provide load balancing across web server farms.

Basically, what load balancing does is make a number of servers providing a web (or other) service look, to the client, as a single server. It does this by providing a virtual IP address or VIP. Client requests are then directed to the VIP, and the load balancing service will direct the client request to the most appropriate server based on a number of differing algorithms (solution dependent).

There are basically three types of load balancing technology:

1. Clustering technology, such as offered by Microsoft Network Load Balancing (NLB).
2. Software running on a ‘dispatcher’ type server which receives the initial request and forwards the request to the appropriate web server.
3. Load balancing switch technology.

Type 2 solutions are very rarely seen nowadays, but were quite common a few years ago. They tend to be unreliable and have very little functionality.

Microsoft’s NLB is packaged within Microsoft Application Center 2000 and provides good load balancing capability for web servers. An add-on from the Microsoft Windows 2000 Resource Kit, Microsoft Cluster Sentinel allows you to make NLB service aware.

Service awareness is an important function of load balancing technologies. Until 4 years ago, most load balancing technologies were server rather than service aware. This meant that they monitored the availability of the server at the network level. Hence, if the network connection to the server was available, it assumed the service running on the server was also available. Sessions could therefore be directed to a server on which the service itself had failed.

Load balancing technologies such as Microsoft NLB (with Microsoft Cluster Sentinel) and load balancing switch technology, now monitor the availability of the service itself, ensuring that session requests are not directed to a server with a failed service.

Typical services monitored by NLB are:

- HTTP
- Chat
- SMTP
- SQL

and with other load balancing technologies this list can be extended to include:

- ICMP
- TCP
- DNS
- POP3
- NNTP
- FTP
- SSL
- WAP
Support for these additional services is dependent on the particular load balancing switch solution. Some switch technologies also allow you to script the *healthcheck*, providing even further functionality.

The most powerful load balancing technology is found within load balancing switches. These devices started as network load balancing switches, providing much the same functionality as Microsoft NLB provides today, but quickly developed to provide support at the application layer, and much of the success of these devices comes from their *content awareness*.

Load balancing switches are able to provide support at layers 5, 6 and 7 of the OSI model and provide support for functions such as URL load balancing (rather than IP or port load balancing); redirection and load balancing services based on the client request itself (for redirection of client requests to cache servers to server static content); and redirection and load balancing services based on the content of a cookie (to identify a particular client or client type, and redirect them to premium services). There are many white papers on this specific technology available from manufactures such as Nortel Networks (the Alteon web switch devices), Cisco, Foundry Networks and F5 Labs.
Summary

This document purposely covers just a few of the technologies required to build a systems infrastructure for e-business, and does not touch on a variety of additional topics such as the management, operations and monitoring of such an environment; or the application and server layer - products such as Microsoft Commerce Server, Application Center 2000, Host Integration Server, Content Management Server, Internet Security and Acceleration Server 2000, Mobile Information Server 2000 or a host of other third party products in this area.

Matters get even more complicated with debates such as the .NET framework vs. Java debate at the application layer, the implementation of component load balancing, and the Apache, Microsoft Internet Information Server or IBM WebSphere debate.

And what about the back-end servers? For example, Microsoft SQL Server, Oracle or Exchange. Or the integration of existing (legacy) host-based systems?

There can be no doubt that building a systems environment to successfully support e-Business is a complex area with a requirement for people who really understand the various technologies, their particular benefits and disadvantages, and how they all fit together.

Unisys have the experience and successful track record, and are able to assist in the design, development and implementation of systems and network infrastructures to provide highly available, highly scalable, highly responsive and highly resilient services necessary for 24x7 operation to support e-Business. We are repeatedly providing these skills and expertise in a structured manner, using the proven SystemFlow methodology. This methodology clearly describes what will be delivered at every stage of the project, and what are Unisys and your responsibilities, as well as the responsibilities of any other parties involved in the implementation of your total systems environment.

For further information contact:

hazel.finney@gb.unisys.com