



There but not really There

Virtualisation overview

With thanks to:

Drs. Q. Laureijs	(qlaureij@redhat.com)	Red Hat
Ing. R. Ivens	(rivens@foundrynet.com)	Foundry Networks
Drs. P.J. Visser	(p.j.visser@i-to-i.nl)	i-to-i
Drs. T. Hatch	(t.hatch@i-to-i.nl)	i-to-i

for providing input, valuable ideas and reviewing of this whitepaper

Also thanks to:

Science and Technology B.V. (<http://www.stcorp.nl>) since I was working there when I started writing this whitepaper

Target audience:

Chief Information/Technology Officer level



Author: Drs. R.B. Gloudemans
Consultant
E-Mail: r.gloudemans@i-to-i.nl
Date: December 31, 2007

Ideas to Interconnect BV
Radex gebouw,
Kluyverweg 2a,
2629HT Delft,
Nederland

Tel: +31 (0)15 268 25 13
Fax: +31 (0)15 268 25 21
E-Mail: info@i-to-i.nl
Web: www.i-to-i.nl

Management summary

Virtualisation is the technique of relating physical computer resources to logical definitions of those same resources, for the purpose of providing a business service with its own optimised, stable and independent IT environment.

There are various technologies and vendor products available to achieve virtualisation, including Xen and VMware. These products comprise a tool called the Hypervisor, whose task is to map logical resources onto physical resources, and a virtual machine manager which organises and controls the logical subgroups of computer resources, known as domains or virtual machines.

Advantages of virtualisation include the need for fewer machines through better utilisation of hardware capacity, thereby reducing hardware, software licence and energy costs. Adjusting air conditioning capacity accordingly will further reduce environmental impact. Because of the virtual nature of the domains, they can be moved with relative ease between machines to facilitate maintenance without interrupting the business service. Performance and demand management are also relatively easy to effect, through adding and removing resources from computer environments as required.

The main areas of concern regarding virtualisation are the inherent security risks, and the greater potential impact of component failure as a result of clustering business services on fewer hardware units. The presence of a virtual machine manager within a hardware system provides a target for attacks on system security. The architecture of the IT environment and organisation needs to be holistically designed and developed in order to eliminate potential weaknesses. Proactive management and maintenance of the IT systems alongside insight in the business relevance of specific components (and impact of component failure) is also essential in the successful deployment of virtualisation.

Virtualisation (especially within the context of a Service Oriented Architecture programme) can contribute to the professionalization of an IT service organisation.

Table of Contents

Management summary.....	2
Introduction.....	3
Technology description.....	5
Para Virtualisation vs. Full Virtualisation.....	6
Virtualisation and Networks.....	7
Other forms of virtualisation.....	7
Availability.....	8
Performance.....	10
CPU.....	10
Network.....	11
Disk.....	11
Recovery after host failure.....	12
Future.....	12
Flexibility.....	13
The Greenhouse Effect.....	14
Security.....	15
Virtual Data Centre.....	15
Guest Separation.....	15
New possibilities.....	16
Ground rules.....	16
Which product to choose.....	17
System Management.....	18
An Example.....	18
A Solution.....	18
Service Management.....	20
Conclusion.....	21

Illustration Index

Hypervisor architecture.....	6
Failover of dom3. A virtual disk can be a file, logical volume or disk partition which is represented by the Hypervisor to the guest as a real disk.	8
Virtual management groups.....	18

Introduction

Imagine giving the order to create a new infrastructure to support a marketing campaign for a new product. Your application developer tells you he needs at least 20 servers for the production and test environments. You, being the technology officer agree with him and tell the IT department to make it so.

The campaign has started and you want to show off the work you've done to the chief. You start in the computer room and ask the operators where the racks with the servers are. They point to a corner of the room where 2 servers and a small SAN are running.....

Fiction? Perhaps, but with modern virtualisation technology not really impossible. Most servers in the computer room are mostly busy doing nothing. Because services have to keep running at all times and because servers are bought with an eye on the load they have to support in 3 to 5 years, computer rooms are filled with surplus CPU cycles, memory and bandwidth. This is the ideal environment to start using virtualisation techniques.

At first this looks like a great idea, but imagine one of those 2 servers going down; this means that a grand total of 10 *guests*¹ will stop working and may have to be restored from backup. There goes the advantage! Or what if one of those guests is hacked, what will this mean to the other guests and the host running them? Can the management organization even support virtualisation?

You won't find the answers to these questions in a typical virtualisation ad. It is a good technique and a very good solution for server consolidation, but the decision to use it should not be taken lightly. The downsides, if not properly managed, might lead to more IT related problems instead of less.

1 *Guest*: A common name for a virtualised system. The term *guest* comes from one of the virtualisation technology vendors. Another large vendor calls them *domains*. Physical systems are called *hosts*.

Technology description

Virtual¹: being such in essence or effect though not formally recognized or admitted

In the world today, virtual systems, also known as guests, means running multiple Operating System instances on one piece of hardware, also known as the host. To all users and applications a guest is indistinguishable from a real system. Virtualisation of systems can be accomplished by a multitude of technologies and a host of vendor products.

The concept of virtualisation is not new². A first reference to it is made by Christopher Strachey in a paper on "Timesharing in Large Fast Computers" in 1959. In the early 60's Atlas created the first, of what later becomes known as, virtual machine. Mid 60's and on, IBM also did a massive amount of research in this area.

The main goal of virtualisation is to provide each application with its own unique, stable, without foreign influences, optimized environment. This goal can also be achieved by giving each application its own server, its own backup server and its own part of the network. The spectrum of virtualisation products varies from technologies like the Java Virtual Machine, where each application gets its own sandbox, to implementations like IBM's hardware partitions.

Somewhere in that spectrum are products like Xen and VMware. These products run on relatively cheap x86 hardware and are able to run multiple guests on one box. These products have advanced rapidly throughout the years; aided by built-in support for these technologies. Even modern desktops have this technology. Software virtualisation has developed to a stage that will very soon be on a par with the hardware solutions of the large Unix vendors.

Virtualisation works by adding a virtual machine manager, also called a *Hypervisor*, to the server. The task of the Hypervisor is to present each guest with its virtual resources and multiplex those virtual resources to the physical resources.

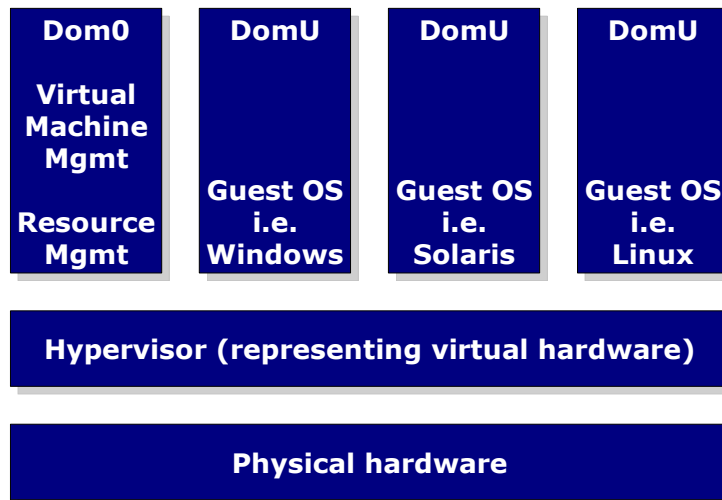
Two main types of virtualisation exist on the x86 platform. Full virtualisation; offered by products like VMware server/ESX, Parallels and Xen, and para-virtualisation; currently only offered by Xen. More on the difference between these two will be explained later.

Xen is currently the rising star under the virtualisation platforms. Therefore, for the rest of this paper Xen will be used as the default example. Other technologies work in the same way and have the same benefits.

Figure 1 shows a diagram of a host with multiple guests. In Xen a guest is also called a *domain*. Two types of domains are present in the figure; dom0 is the management domain. Other domains initiated by dom0. Dom0 has the highest privileges possible for a domain. Dom0 can control the resources the other domains are using. There can be only one dom0 per physical host. The other

1 source: Merriam-Webster Online dictionary

2 <http://www.kernelthread.com/publications/virtualization/>



domains are called domU for "User domain". A domU domain is isolated from the other domains; it can only access its own virtual resources. A domU is by default prohibited from accessing dom0 or any other domU.

In this paper the term "guest" is the same as a "domU domain", unless dom0 is specifically mentioned.

Figure 1: Hypervisor architecture

Para Virtualisation vs. Full Virtualisation

If full virtualisation is used, the operating system operating in a domU does not know it is virtualised. Therefore all operating systems can operate in the domU domain unmodified. This however sets a huge task for the Hypervisor as all virtual hardware needs to be represented as real hardware to the domU. New processors have extensions to help reduce this overhead. Expect a performance loss of around 5% or more, compared to a real system. The actual performance penalty depends on the application running in the domU.

A para-virtualised domU knows it is virtualised. It also knows of the existence of other domains and can even communicate with them. The big disadvantage here is that the operating system needs to be modified to run in a para-virtualised domain. Because hardware representation in a para-virtualised world is much simpler, a para-virtualised domain can run with a performance penalty of less than 5%.

The fact that the domUs are not fully separated also looks like a security risk, but all domains including dom0 are still prohibited from accessing hardware directly and inter-domain communication is by default limited to communicating about resource usage. From the application perspective there is no difference between the two virtualisation types.

From the device driver perspective, both virtualisation technologies have a major advantage; the representation of hardware to the guest always looks the same, regardless of the real physical hardware. This creates an extra degree of freedom regarding application and hardware support matrices. If for example an application is only supported on version A of the OS and the hardware drivers for the server it needs to run on are only supported on version B, then it is possible to run version A in a domU on a host where dom0 (which has to be compatible with the hardware drivers) is running version B. This way, no support matrix is violated.

Virtualisation and Networks

From the network point of view, virtualisation has been common practice for decades. Almost every physical network is divided into multiple logical networks using Virtual LANs (VLAN). In a normal situation, one network port on a network device is only member of one VLAN. This allows the network department to be fully responsible for network separation.

To create an optimal flexibility regarding which guest is running where, the VLANs should be terminated at the host port instead of the network device port. In other words, a trunk is created to the host. Guests, other than dom0, need not be aware of the trunk. There is simply one network interface per VLAN as in traditional configurations.

Other forms of virtualisation

This paper only handles virtualisation on system level. There are many other forms of virtualisation. Network level is one example, but virtualisation also exists at storage level and application level. Many large companies utilizing large SAN and/or NAS enclosures are already using storage virtualisation on a large scale.

Availability

Having fewer systems means that the chance of a piece of hardware being defect at any given moment is reduced. But if a hardware defect does occur, multiple guests are affected and the impact on the service portfolio is likely to be larger than in a non-virtualised environment.

To cope with this, technologies exist to initiate the same guest on another host, synchronize the guest and then disable the original guest. This is in effect the transfer of an operational domain to another host. In the real world this process takes a few milliseconds, which is a figure that cannot be met by any Operating System level clustering software.

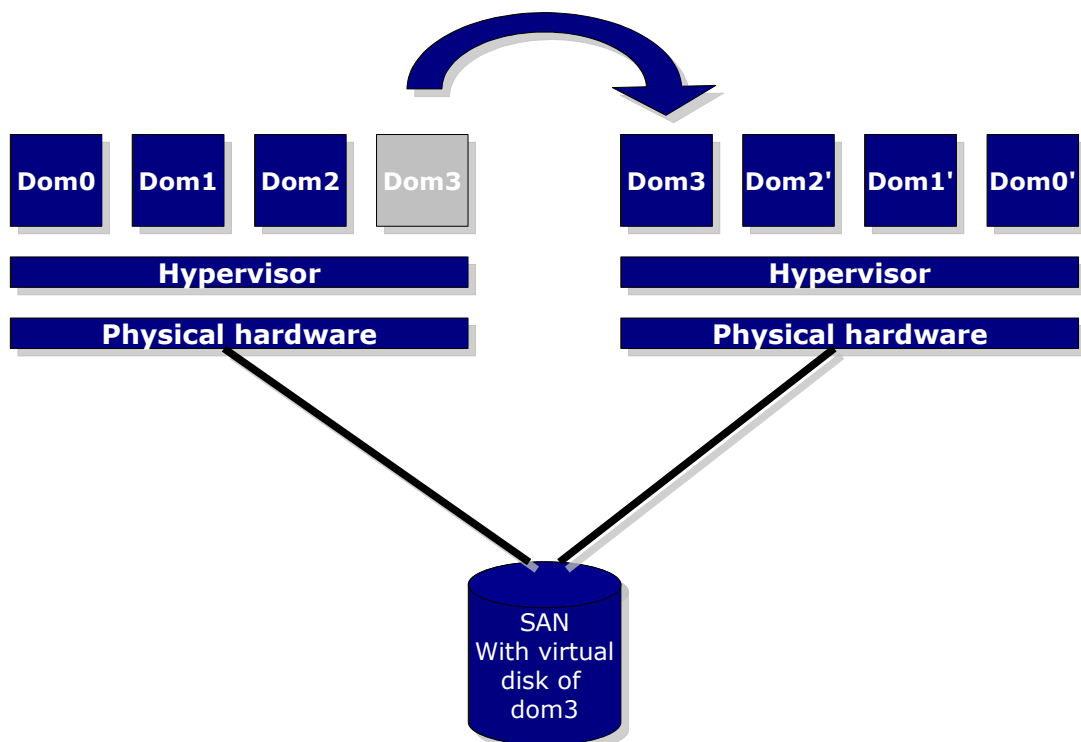


Figure 2: Failover of dom3. A virtual disk can be a file, logical volume or disk partition which is represented by the Hypervisor to the guest as a real disk.

The ability to move guests from one host to another does place some constraints on the infrastructure. Guests use *virtual disks*. A *virtual disk* can be a file, logical volume or disk partition which is represented by the Hypervisor to the guest as a real disk. The virtual disk of a guest must be available on multiple hosts. If the virtual disk is located on a SAN or NAS, it is possible to make the virtual disk available on two or more physical systems.

For optimal availability, hosts should be clustered. Traditionally Operating System clusters used to move applications from one server to another. In this time and age it is about moving guests between the hosts. If a host is taken off-line at a predefined moment, for example for maintenance or a non-critical defect in the hardware, virtualisation will offer much better performance than the Operating System level cluster.

However, in case of a critical defect, hosts will go off-line with no warning. The recovery time from such an event depends on the infrastructure configuration. If for example a host is standing by, the recovery time will be much longer than when the guests are distributed over the complete infrastructure. Ten guests booting at once will create a huge host load, whilst one guest booting on a host which has already 10 guests running will have a much smaller impact. More on this in the section on performance.

Some products can cope with this case of failure; a copy of the guest is running on another host at all times and kept synchronized. This way, when the active guest goes down, the guest standing by can take over without pause. The downside of this configuration is a high and continuous demand for resources to run a single guest, including substantial network traffic.

Should a guest go down because of an application change that failed, or because of a hack, recovery is straightforward. Most virtualisation products support some kind of *snapshot* technique. In a *snapshot* the current contents of the guest are duplicated. This can be done on-line and is only limited by the amount of storage available. The recovery of a guest is as simple as copying one file. The snapshot can be used on other hosts, irrespective of the hardware configuration of the host, as long as the processor type does not change.

A small note on security is required here. Synchronizing guests between hosts involves the transmission of raw memory data over the network. Therefore this communication should always take place via a separate network, or through an encrypted tunnel.

Performance

Managing the performance of a guest is a difficult subject. How many guests fit on one host? One way to find out would be to look at the statistics of the host and add the expected needs of the applications and just keep adding until the host is full.

In practice this will work, but to get the full benefit of virtualisation two new time-related parameters should be taken into consideration:

1. Real hardware is bought to the specifications that the application will have at the end of the life cycle for the hardware. In a virtualised environment, a guest can be provided with what the application needs initially. This is often a lot less. If the application requires more resources, they are just a click away.
2. The load profile of any application will vary during the day. Batch-oriented applications will generate a heavier load at night, while on-line applications will have a bigger load during daytime. This should be factored in when determining which guests should run where.

On the other hand, why bother to do the exact calculation? As long as there is some slack in the infrastructure a running guest can always be moved.

Technically it is that easy, but financially it isn't. Total cost and performance of the infrastructure should be predictable. This performance requirement is probably also formalized in a service level agreement. The chance of predicting the required resources accurately is lower. This increases the financial risk. If an organization is not willing to take this risk, much of the gain of virtualisation might be lost.

While this sounds serious, it need not be that bad. Statistically, underestimating happens as often as overestimating. So with a couple of guests on one host, the estimation errors even out. If this is not the case, performance will be affected and one or more guests will need to be transferred to another host.

CPU

On a multi-CPU host with multiple computing cores per CPU, a discrete number of cores can be assigned to a guest. This makes CPU performance predictable because there is no dependency on the load from other guests. This does have a negative impact on total power consumption though. More on this in the "Greenhouse Effect" chapter.

The performance per CPU can be seriously hampered if the CPU cannot reach the system memory unhampered. This is especially the case in older Intel Pentium/Xeon systems, where there is one overcommitted data bus to which memory and CPU are connected. In more modern systems this data bus is no longer overcommitted, but with CPU speeds increasing in new CPU models this might be the case again.

Network

When running a substantial number of guests on a host it will not be possible to give each guest its own network interface. Guests will have to share physical interfaces. Most guests never use more than 10% of a 1 Gigabit per second (1Gb/s) network connection, excluding backup over the network. To safeguard network performance intelligent switches provide a solution. Many advanced switches provide support for quality of service and/or rate limiting policies. These can be used to limit the amount of traffic that each guest may generate. A 1Gb/s network interface could be partitioned into ten 100 Megabit per second (100Mb/s) interfaces. This is still enough for most applications.

Well before the end of this decade we will also see the breakthrough of 10Gb/s networking. Whilst these networks are mostly fibre-based at the moment, 10Gb/s over copper will become more common. Switching hardware for such a network is still scarce, but this will change soon. One such interface will provide enough capacity for many guests.

While this sounds great, there is a catch. 10Gb is more than a host can handle. If a 10Gb/s card was built like regular interfaces, the maximum performance would be no more than 5Gb/s, causing heavy CPU load at the same time. That is why cards for these speeds must be equipped with their own processor to offload work from the CPU. This enables usage of the full bandwidth. Not all 10Gb/s cards are equipped with such a processor, so this is a point of attention when buying hardware.

But still, adding more than two such interfaces to an x86 system, no matter how many CPU cores it has, creates problems. Depending on the actual system, the network card can saturate the internal data bus in the host. This then creates other problems like slow access to, for example, disk devices.

From the host point of view, it is easy to provide enough network capacity. One 10Gb/s interface provides enough bandwidth for 10 to 100 guests. On the network level this is a different story. Certain large network manufacturers have been building oversubscribed switching equipment for years. Many companies rely on these brands for at least the core of their network. What this means is that for any given switch, not all ports can run at full capacity simultaneously. This is a major factor when (re-)designing the infrastructure.

Disk

The disk is the biggest bottleneck of all. Compared to CPUs, system memory and even the network, disk performance is very slow. That is why each performance critical guest should have a one-to-one mapping of virtual to physical disk. This will guarantee disk performance even when other domains are causing heavy disk activity.

The next bottleneck is access to the disk. Most internal I/O buses are fast enough to allow unhampered access to the disks. However, for availability purposes SAN or NAS storage is mostly used. The average performance of any physical hard disk is around 20-30 Megabyte per second (MB/s). This translates to roughly 300 Mb/s in terms of network speed. Peek speeds are even faster since physical disks

also have built-in cache. An improvement to 200-300MB/s for cached data is not uncommon. When SAN/NAS solutions are deployed, with load spreading across several disks, performance for sustained reads and writes should be somewhere between these speeds. This means that an I/O heavy application like a database would at least need a 1Gb/s connection to the physical disks' location.

A typical Host Bus Adaptor (HBA) used to access remote storage has a bandwidth of 4Gb/s. If the disk I/O bus may not be overcommitted, 4 guests can be allowed per HBA, based on the calculations from the previous paragraph. It is possible to use multiple HBAs in one host. The load can then be load-balanced over the HBAs and if one fails, the remaining cards can take over. Based on the actual load on the host at the moment of HBA failure, performance might degrade. This technique is called multi-pathing.

Recovery after host failure

As noted before, having a host on standby is one of the typical pitfalls in a virtualised data centre. Starting up guests always goes hand-in-hand with a high I/O load on the disk. The guests will be delaying each other's boot.

The best way to do this is to give each host in the infrastructure some overhead. When one host dies, its guests should be distributed over the infrastructure, where no two guests go to the same host. This will allow for the quickest recovery.

If the software that controls the host clusters can be made to interface with the (performance) monitoring software another possibility opens up; least impact automatic fail-over. Each host will have the pre-defined pool of resources available to cope with disaster recovery. The size of this pool will vary per host. The causes for this are diverse due to predicted loads always differing from the actual loads, due to increased (periodic) business demand.

When the cluster knows which hosts have the lowest load, it can migrate the guests to those nodes. This way recovery does not have an adverse impact on the service levels. A good planning of where domains should run will also accomplish this, but at a higher management cost.

Future

The one weak point of virtualisation on the x86 platform is resource isolation. When one domain starts claiming more resources than anticipated the performance of the other domains will be impaired.

This has not gone unnoticed. Several efforts are under way to create mechanisms that will provide a given set of virtual hardware with a given performance. The latest Xen versions are already able to prevent one domain from overloading the host's CPU. Additionally various parties are undertaking research to extend the resource separation even further. One such mechanism is ShareGuard¹, which is expected to be incorporated into Xen in the near future.

1 http://www.hpl.hp.com/personal/Lucy_Cherkasova/papers/xenqos-mware-camera-ready.pdf

Flexibility

Alongside performance management, infrastructure flexibility is one of the key items in reducing the Total Cost of Ownership (TCO) of the infrastructure. Virtualisation facilitates the flexibility in various ways.

The first is through the abstraction of the hardware layer. All guests are provided with the same “virtual” hardware, regardless of the physical hardware underneath. This not only allows guests to switch to any host, but also limits the system management efforts. Many servers require drivers that are not delivered with the operating system and need special attention when for example patching or upgrading OS kernels. With virtualisation the total number of machines in the infrastructure is less than without virtualisation. There is therefore a smaller number of servers with the non-standard drivers.

The fact that there are less physical systems also improves flexibility within the data centre. Less space is needed and less power and cooling resources are used.

Provisioning is also easier. All hosts will look exactly the same. So a single installation image is sufficient. Guests can be created by copying virtual disks. There is almost no difference between provisioning one web server or a hundred. System configurations and patch levels tend to change with time. If an extra web server is needed, the current server is copied and given a new network address. All configuration changes and patches are already applied. Restoring a guest is also easier; just copy back the snapshot.

The Greenhouse Effect

Virtualisation decreases power usage. But is this true? There are indeed fewer servers, but each server is working harder, thus consuming more power.

The key here is efficiency. Take, for instance, one of the new Quad Core CPUs from either Intel or AMD. The biggest processors have a maximum power usage of 120 Watts (120W). But when these processors are idle, they still consume about 30W.¹

Power supply efficiency and the efficiency of all other components increases with the system load. The difference in power usage between an idle server and a fully loaded server is somewhere in the region of 50%². So by consolidating servers using virtualisation, a substantial amount of power reduction is achieved.

Even if the total amount of CPUs is kept constant, when two single CPU servers are consolidated into one dual CPU host, less power is used. This is because there are still shared components in the server. Efficiency of power supplies is somewhere in the 80-90% region, with the 80% for the idle server³, although the gains are more significant when the number of CPUs in the data centre is reduced.

Less power usage means less cooling is required. To maximise efficient use of air conditioning units either fully utilize data centre space or, if cooling needs are substantially reduced, shut down one or more air conditioning units to keep the remainder running at optimum efficiency.

Not only does the corporation gain from this technology, but society in general gains due to the reduced emission of greenhouse gases. Note that, because the total number of systems is smaller, the environment also gains from not manufacturing the servers that are not needed.

1 http://www.amd.com/us-en/assets/content_type/DownloadableAssets/43761A_ACP_WPv7.pdf
2 <http://techreport.com/articles.x/7429/15>
3 <http://www.anandtech.com/showdoc.aspx?i=3040&p=11>

Security

Security is often overlooked when using virtualisation. A couple of new issues need to be addressed to prevent an increase in infrastructure risks. Security has two major points of attention:

1. Everyone with access to the host or dom0 domain has access to all guests running on that host. This can be compared to having access to the data centre. Not everyone is allowed to enter the computing floor; the same principle should apply to dom0.
2. Servers are separated. Inter-server connections need to go over the network or via a SAN disk. Guests run on the same hardware, what about separation here?

Virtual Data Centre

Coming back to the first point, if network trunks are terminated at the host level instead of the switch level, responsibility for the network also shifts. Previously the network operators determined which server had connections and who was allowed to access which VLAN. This is no longer true. Part of this responsibility has shifted to the staff operating the host and dom0 domain.

Additionally measures should be taken to prevent abuse of dom0 domain. Dom0 should have a minimal set of software which runs in a properly configured environment. Operators rarely log in on dom0 domains except to create, remove and restart other domains. If the hardware contains an Out of Band management board, with which the console of that server can be accessed, one might even consider running no services except the virtual machine monitor. Even typical services like secure shell access are not needed.

Guest Separation

The second point “server/guest separation” is also not trivial. In theory, in a fully virtualised environment, guests cannot have any interaction. Guests are not able to access the Hypervisor. In reality this is not true¹.

Virtual hardware is simulated by the Hypervisor. This Hypervisor is a process that runs with administrative privileges. Flaws in the code which takes care of simulation of those virtual devices could be exploited from the domains. This could enable a malicious person to gain control over the Hypervisor and thus administrative privileges on the hardware. From there, this person could access the other domains.

In this respect para-virtualisation is much better. The guests are not fully separated but the design of the architecture is much more secure. Elaborate emulation of hardware devices is not necessary. Instead each guest including dom0 has a front end driver which matches to the real hardware drivers.

In practice the effect of a flaw in the driver code is limited. For example, if the flaw is in the network driver code, the attacker needs to get a specially constructed network packet to the server. To start with, the attacker needs to

1 <http://taviso.decsystem.org/virtsec.pdf>

know which virtualisation product is being used. Then he needs to make sure that his constructed network packet gets to the guest in question. This is not as easy as it sounds. The chance is large that this specially crafted packet does not conform to the official network standards, because all vendors test their products using these standards. Routers between the attacker and the guest will then discard the packet. But if the infrastructure is a high profile target, this is something to take into consideration.

Though para-virtualised guests are not fully isolated, from the application point of view, they are. In the future this might change with technologies like sHype¹. sHype is a security architecture for virtualisation environments that controls the sharing of resources among the domains according to formal security policies. Communication between the guests may go beyond the network, for example via direct memory-to-memory copies.

New possibilities

sHype creates new possibilities for applications. For example within a Service Oriented Architecture (SOA), each service may be available in the form of a domain running on a machine. In such an environment sHype could be part of the common data bus which connects the individual services. This would increase performance and control.

Further considering the security aspect, there is another advantage of this technology. Using sHype for SOA would mean a step away from the centralized Message Brokers. The Message Broker is essential in SOA architectures and all services rely on it. If the Message Broker is compromised, all services are compromised. Thus using sHype for SOA reduces risk.

Ground rules

If virtualisation is to be used in environments where security is very important, the infrastructure needs to be well conceived:

- Avoid mixing application tiers on one host
- Make sure responsibility for network access does not strand between departments
- Severely restrict access to dom0
 - Do not allow access from domU to dom0
 - Do not install any additional services on dom0
 - Allow only a subset of the most trusted system administrators to login to and maintain the hosts. Allowing only the most experienced administrators also limits the risk of bringing down all domains because of human error.
 - Use authentication tokens for dom0
- Use auditing services

1 <http://www.acsac.org/2005/papers/171.pdf>

Which product to choose

There are several vendors active within the virtualisation market, each with their own products with their respective strengths and weaknesses. Which product is the best choice?

To answer this question, there are two aspects to consider; the Hypervisor and the management tools within each product.

The Hypervisors are almost irrelevant when choosing the product. The performance and capabilities of the various Hypervisors are rapidly converging. One performs better with application A, another with B etc.

However, the management tools do vary considerably between products. Windows organizations tend to prefer VMware's and XenSource's management interfaces best, Unix organizations favour the command prompt of the Open Source Xen/KVM/etc., although graphical management interfaces are emerging here too.

The best choice for the long term would be a product that does not create a vendor lock-in. This is possible nowadays. Libvirt¹ is an abstraction layer for virtualisation. All mayor players in the virtualisation arena have, or are working on, libvirt compatibility.

In practice, this means that one can boot a guest with the Hypervisor of the day. Now it is possible to choose the Hypervisor which performs best with the application, without system management implications. The only boundary condition for the moment is that Hypervisors don't mix on a host.

The organization can choose the management tool they like best, as long as it is compatible with libvirt.

1 <http://www.libvirt.org>

System Management

Using one virtualisation strategy across the company will optimize infrastructure flexibility. While this sounds trivial, making this a reality is not easy. This chapter describes a possible problem and solution to show that organisational changes could be required to get the best result from virtualisation technology.

An Example

Think of an organization with a large multi-platform infrastructure. Such organizations typically have separate organizational units to manage each platform and one for network management.

If virtualisation is only used within the units, then all hosts managed by that unit will need more overhead to recover guests from a failed host, as there are fewer hosts available to distribute the guests across. This increases the Total Cost of Ownership of the infrastructure.

Moreover, disagreements between units may lead to even more cost increases. If for instance the network unit cannot agree with a systems management unit, trunking to the host will be impossible. This will create boundaries within the infrastructure managed by one unit.

For holistic management of the infrastructure the separate units will have to work together. Take management of the dom0 domains for instance. One unit will have to take responsibility for them and the other units will need to trust this unit. In theory this is not a problem, but in the real world where different units are involved in hosting a service, this may not be the case.

A Solution

Since the dom0 domains are high impact environments only the most experienced and trustworthy staff should maintain these domains. These people should also understand networking and

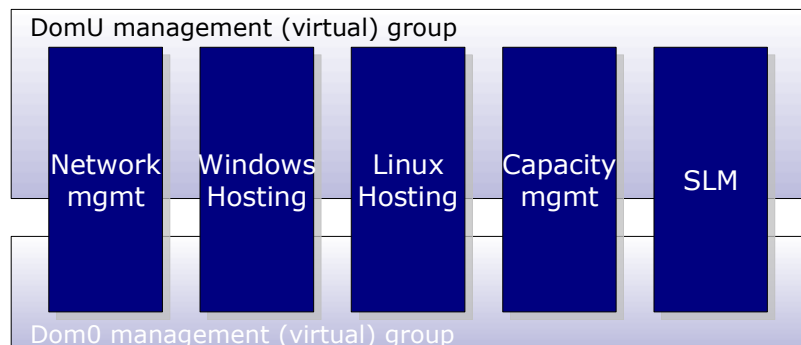


Figure 3: Virtual management groups

have at least basic knowledge about the operating systems and applications that are running in the domU domains. A possible solution would be to create a new unit consisting of the most experienced people of the current units and make them responsible for the Hypervisors and dom0 domains. These people can remain within their existing units, since managing the dom0 domains should not entail much extra work; it is simply an additional role. In effect a virtual unit is created.



There but not really There – virtualisation overview

Drs. R.B. Gloudemans
Consultant

Because the virtual unit is made up of people from all existing units, the trust issue should be less prominent. This will make it easier to agree on and exploit one strategy.

Service Management

In a non-virtualised environment there often exists a one-to-one relationship between applications and servers. Performance is not usually an issue and availability is relatively simple: a server is up or a server is down.

In a virtualised environment it is not as simple. Applications share hosts and influence each other's performance. When the infrastructure fails, choices have to be made. It is no longer a question of 'up or down': when one host breaks down the systems manager will have to reallocate the remaining resources. He or she will have to decide which services get priority over others. This not only requires technical knowledge but also knowledge of the business processes and the way the services contribute to the business.

Alternatively, service priorities can be recorded in the Service Level Agreements (SLA). The systems management unit then needs to make sure that priorities are not clustered on one host, preventing too many priority conflicts during service recovery.

This approach to services changes the scope of Configuration Management. In the traditional Configuration Management Database (CMDB) we typically keep information on hard- and software. By adding virtual components, applications, services and their mutual relations, the CMDB will provide us with information with which to make appropriate decisions.

Changes in a virtualised infrastructure often have impact on several services and when a change fails the business processes can be seriously affected. A well functioning Change Management process therefore becomes even more important. Virtualisation increases the amount of people involved in changes. For example, when servers have to be patched, a short period of downtime is inevitable. In pre-virtualisation times one would agree with the application manager on a suitable moment to bring down the application. In a virtualized environment there are many different parties involved, each with their own interests, and reaching a mutually agreed change time slot might be difficult and require, for each change, a serious amount of negotiating. It might be a solution to agree on a recurring service window in which the infrastructure is available for maintenance. It is self evident that this service window should be included in the SLA with the clients.

From a Service Management point of view, virtualisation has another huge benefit: it requires the IT department to change its focus from technology to the business processes, thereby bringing true IT-Business alignment one step closer.

Conclusion

There but not really There; the title fits virtualisation in more than one way. The technology is maturing fast, but lacking in a few key areas at the moment.

In the security area, most problems are caused by the desire to put the new products into production too fast, coupled with the fact that no-one has looked at the security of virtualised environments closely enough yet. This should be easy to correct by the software vendors and communities that build these products. However, virtualisation products are safe enough to run the majority of IT services now, if current shortcomings are taken into account when designing the infrastructure.

A lot of SLAs tend to focus on global metrics based on system statistics, like CPU load. Not all of these statistics can be propagated easily to guests. It could be done, but it is better to re-evaluate the SLA and choose service-centric metrics instead of server-centric.

Virtualisation holds many promises. If implemented properly, availability and performance can be increased and hardware, licence and energy costs decreased. There can be a huge increase in infrastructure flexibility. But to take full advantage of these benefits, organizations and working practices need to change. Without proper design and appropriate organizational changes, benefits could turn into costs.

The basis of a successful deployment of virtualisation technology is a knowledgeable IT staff. If the knowledge level of the IT staff is seen as a problem, do not use virtualisation. On the other hand, if the staff is truly capable, the technology that is there but not really there should get you there.



There but not really There – virtualisation overview

Drs. R.B. Gloudemans
Consultant

Ideas to Interconnect BV (i-to-i) is a services and solutions company that helps customers exploit information and communications technology for business advantage. We are consistently at the leading edge of change – working in partnership with our customers to create added value by implementing “best-working-practices” to increase revenue, decrease costs, and increase the quality of service. Normally working with our customers on key strategic initiatives, we deliver industry knowledge and technical expertise with “back-to-basics” common sense innovation. We believe in doing business ethically – building trust and a strong, long-term relationship with our customers. We also believe in sharing risks with our customers - making their success our success.

Ideas to Interconnect BV

Radex building, Kluyverweg 2a, 2629HT Delft, The Netherlands

Tel: +31 (0)15 268 25 13, Fax: +31 (0)15 268 25 21

E-Mail: info@i-to-i.nl, Web: www.i-to-i.nl