

Reducing Complexity in Network Infrastructure

Automation in the Network Infrastructure for Agility at Scale

By *Andreas M. Antonopoulos*
SVP and Founding Partner, Nemertes Research

Executive Summary

Network infrastructure services, such as DNS, DHCP and IP-address management (also called DDI), have remained mostly unchanged for more than a decade. These network services are now facing challenges from new technologies such as virtualization, VOIP, and video. Meanwhile, business trends such as globalization, the emergence of the virtual workplace, and mobility are leading to the proliferation of connected devices and an explosion in the address space, putting further pressure on DDI services. Overall, these changes are making the management of network services simultaneously increasingly critical to network operation, efficiency and stability, and harder to achieve. As networks continue to grow in size, importance, and complexity, organizations need to implement network services that are secure, scalable and fault tolerant.

More Devices in More Places

Networks are growing in size and complexity. Network addresses are proliferating because enterprise IP networks are expanding to include an ever-wider set of device types: no longer just computers, but also phones, building control systems, surveillance systems, door controls, sensors, and heavy machinery. The transition from an Internet of computers to an Internet of people and things means an explosion of connected devices, each with its own unique IP address.

Another factor in the growing size and complexity of networks and their address spaces is the combination of mobility and globalization. Enterprises are more distributed than ever before, with a growing percentage of staff operating across multiple offices, regions, or countries. Nearly 90% of Nemertes research participants say they operate a “virtual” organization in which members of distributed workgroups must collaborate with each other across multiple locations, as well as with partners, suppliers, and customers. Virtual workplaces include branch offices, home offices, hotels, and airports. As companies increasingly

operate globally—as both large and small organizations do today---this trend will only increase.

Branch office scope is broadening to include the “micro-branch,” mobile worker, or single-user telecommuter site. An overwhelming majority of companies (85.6%) increased the number of telecommuters in 2009, after two years of relatively mild growth in the number of telecommuters (17% in 2007 and 20% in 2008).

The Address Explosion(s): My Own Private Internet

All of this is leading to an explosion in the number of IP addresses. Because IP addresses represent network identity and location, every single device in existence may need multiple IP addresses as it hops from network to network. So more devices (identities), multiplied by more locations and increased mobility brings us an explosion of IP addresses.

Another trend leading to an explosion of addresses is machine-to-machine communications. Applications that used to be “monolithic,” housed on a single server, are now distributed among multiple loosely coupled modules that communicate with each other over IP. In other words, monolithic applications became three-tier applications, and are now becoming n-tier service-oriented composite applications. Components that used to communicate using inter-process communications (IPC) in a monolithic stack are now running separately and using TCP/IP communications. If they are distributed across multiple physical servers or virtual machines, each now uses an IP address.

As shown in Figure 1 below, the number of Internet-connected devices jumped quite dramatically, doubling in size between 2004 and 2005. Since, constraints on public address space have kept that growth from accelerating, though it has continued increasing at a steep angle. In 2010, Internet-connected devices were expected to pass 5 billion. Behind the scenes, however, another picture is unfolding: non-public address space is growing even faster. Since public addresses are scarce (IPv4 space is running out) and machine-to-machine communications are increasing, a lot of the address growth is behind the corporate perimeter. Corporate networks are growing at a faster pace than the Internet, because they do not face as tight constraints on address space.

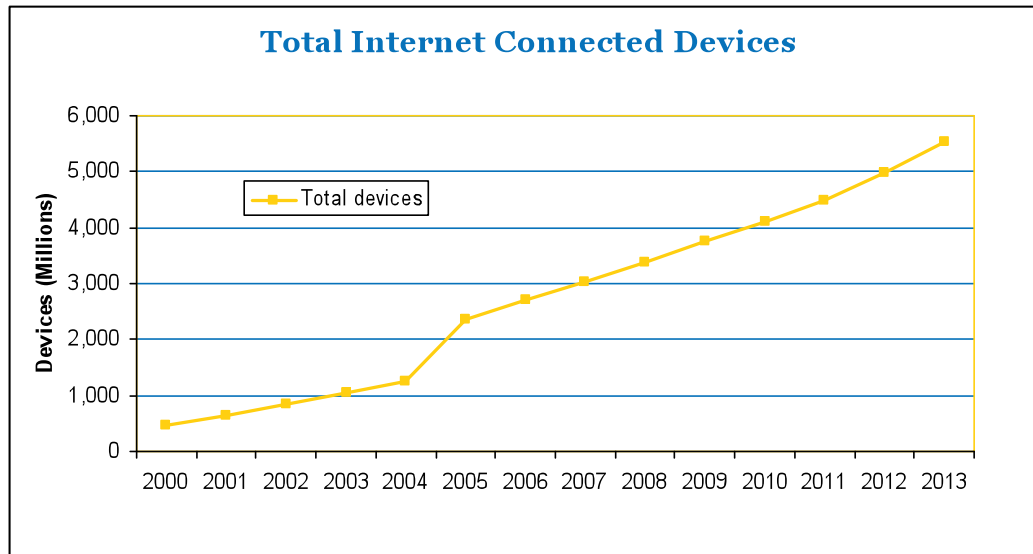


Figure 1: Total Internet Connected Devices

Virtualization and cloud computing further complicate the picture. Virtual machines encourage modularity by making it easier to deploy application modules in separate virtual containers and manage them independently. The virtual machines each have a virtual MAC and IP address and are connected to each other by virtual Ethernet switches and bridges. The result is a huge number of IP addresses, even within a single physical server.

IP version 6 (IPv6) introduces two significant changes in corporate networks. Firstly, it requires new protocol stacks in all the routing and switching equipment. Secondly, it makes the address space enormous.

The adoption of IPv6 will require changes to the DNS and DHCP infrastructure. These changes are in the early stages of deployment in the Asia-Pacific Rim countries and in U.S. federal government agencies such as the Department of Defense. Before an organization can deploy IPv6, all supportive infrastructure needs to be able to handle the IPv6 address space.

Comparing IPv4 to IPv6 in terms of address space is difficult because the human mind cannot easily grasp the numbers defined in exponential terms. In IPv4, there are 32-bits for the IP address, creating a total addressable space of 4.3 billion. Now this may seem like a lot, but on a planet inhabited by more than 6 billion people, it imposes certain limitations. The usable space is further reduced by artificial subdivisions, leading to the broad adoption of Network Address Translation in order to conserve IP addresses. IPv6 changes all of this by allocating 128 bits for the address, allowing for 3×10^{38} total IP addresses. And this is where the human brain balks: 10 followed by 38 zeros is a number with few physical equivalents outside the domain of cosmology. To provide a concrete example consider that with IPv6, each human on the planet could have an entire Internet's worth of IP addresses.

Although IPv6 includes some features that make address management simpler, the fact remains that it vastly increases the volume of addresses needing management.

No More Infrastructure as Usual

Outside the issue of IPv6 adoption, we're seeing the proliferation of "non-computing" IP-addressable devices in the enterprise. IP telephony, for instance, has introduced a new class of device into our newly converged networks. Not only are IP "hard" phones proliferating in many enterprises (because of substantial ROI, as documented in Nemertes Research benchmarks on convergence), but VOIP is also spreading in other forms. We see more than 75% of organizations deploying soft-phones, and an increasing number using real-time communication dashboards, which combine instant messaging, email, and voice. Also, we see the emergence of multi-network cell-phones with VOIP over WiFi adding to the IP address space. Beyond VOIP, we also see the rapid adoption of streaming video, video conferencing and telepresence as further drivers of network complexity and device proliferation.

Wireless devices provide yet another challenge. Not only do wireless devices add to the number of addressable devices, but they create a highly dynamic and variable address space as they flit into and out of coverage areas. Allocation of IP addresses is no longer determined only by the length of a DHCP lease, for a static and always-on device like a desktop, server or stationary laptop. Instead, the device movements in and out of coverage determine the time and duration of the IP address assignment requiring much more sophisticated and automated address management.

PDAs and smartphones combine the problems of mobility and address proliferation: even more devices to manage and more ad-hoc connections in the wireless environment. These devices make secure auditing of the IP address space imperative and more difficult.

Today's enterprise environment is still in the early stages of the wireless and mobility revolution. Ubiquitous IP accessibility is still a far-off goal, but the basic elements are coming together: Multi-modal and software radios that can jump between frequencies and protocols for 3G, 4G, WiFi, 802.11n, and Bluetooth make it more likely that all devices will be IP-accessible and -addressable in the near future. Convergence between voice, video and data in the content domain and cellular, wireless and wired in the medium ensure that IP addressability will become the norm for all kinds of devices we cannot imagine today.

Meanwhile, outside the domain of computing and communications devices, we are seeing IP addresses pop up in unexpected places. Building management systems, security camera systems, heating, cooling, lights, vending machines, and elevator management are just some of the technologies that are becoming networked with IP. Several companies are working on bringing bricks-and-mortar management into the network space through the use of IP-enabled sensors and

actuators. Whether the end goal is increased security or better energy efficiency for buildings, more and more devices are coming “online” and need IP addresses.

The bottom line is that the number of IP-addressable devices continues to skyrocket. The question then becomes: How do you manage all of these IP addresses? You have to register, allocate, monitor, control, distribute, and renew with security, performance and reliability at least as good as today’s. Clearly, our current tools and practices need to be re-vamped.

Network Complexity Leads to Business Instability

Scale isn’t the only challenge, either. With increased complexity comes decreased stability. Complexity leads to a greater chance of manual errors, especially in an area of networking where the vast majority of configuration and management is still done manually. In networking, as in many technical disciplines, human error is at the root of many problems. Seasoned network engineers use the expression “fingers on keyboards,” when identifying the cause of a problem to be human error. Beyond the human factor, complexity leads to more instability because of the increasing number of relationships and interactions between components. As the number of connected devices and addresses increase, the number of possible interactions between them increases at an exponentially greater rate.

Complexity makes systems uniquely susceptible to change: When rapid change is applied to a highly complex network, the possibility of error increases even further. As more devices are connected to the network, the rate of change in system configurations increases accordingly. Each change in each system can have unintended consequences on other systems, further increasing the chance of a problem or instability.

Adding to the problem of complexity, today’s networks also suffer from a lack of visibility. Many network managers honestly admit they have no idea what’s happening in their networks. And virtualization only exacerbates this problem. The ubiquitous adoption of server virtualization in five years has resulted in a massive change within corporate data centers. There are now two distinct layer-2 networks: physical and virtual. The virtual layer-2 network resides inside virtualized servers and connects virtual machines. This new network is in many cases almost invisible to the traditional management tools of the network staff. Most of it is managed through proprietary tools like VMWare’s VirtualCenter.

A significant trend transforming the data center is the flattening of layer-2 networks. In a traditional data center, we find three distinct tiers of switched network: core, distribution, and access. In modern, virtualized data centers we see flatter networks with only two layers of switching (core and access). Flatter networks are easier to deploy and remove some management problems, but they also make for much larger subnets. As the layer-2 network becomes larger and less segmented, problems in layer-2 can propagate much further and cause much broader instability.

A final challenge to network stability comes from the phenomenon of amplification. A small problem in the lowest layer of a network will cause instability that is amplified up the layers, as more and more dependent services and applications are affected. Instability in a network service, for example DHCP, can have an amplified effect as it causes instability in all the servers that depend on it. Problems like this are very difficult to troubleshoot because they manifest themselves at higher layers in a way that cannot always be related to the underlying cause.

Network Services Need Automation and Re-Architecture

Despite all this, many organizations still approach IP-address management as a somewhat secondary consideration. Since the basic tools are available for free on any platform, many companies deploy DNS and DHCP servers on an ad-hoc basis with little or no central management. Various software packages are used to serve DNS, but the most common are the Berkeley Internet Naming Domain (BIND) developed in the late 1980s, and Microsoft's Active Directory system. These tools are efficient but not very user friendly, and they require skilled administration. More importantly however, BIND and AD only support the most basic aspects of IP address management, effectively just providing DHCP and DNS for a single, small domain, with only a few hundred addresses. Beyond a small number of IP addresses, BIND and AD start showing their limitations as the address space becomes unmanageable by the network operations staff. Keeping track of the IP address registration, allocation, host association are all things that have to be done outside BIND and AD.

How do companies manage the IP space served by these tools? Surprisingly enough, in many organizations IP-address management consists of a paper record or a spreadsheet. And while these manual approaches might work for a domain containing a few dozen static hosts, they cannot cope with the two major trends in IP-address management: rapid growth of the address space and the highly dynamic nature of transient, wireless and mobile devices. For enterprises and service providers today, large IP address space and roaming/mobile users are not the exception: they are the norm. As a result, many organizations are struggling to keep up with antiquated IP-address management practices that have been imposed on IT by the selection of "free" tools as the only means of IP-address management.

Today's network service management tools take a holistic approach to address management. Beyond serving a single IP address to a host, they offer the ability to manage the entire IP address space of an organization from a proactive capacity planning perspective. Managing the IP space as a *network resource allocation* problem, gives IT managers and CIOs a whole different perspective on address management. Instead of looking at address management problem as basic "housekeeping," it becomes a powerful tool for security, capacity planning, availability and growth management. IP-address management (IPAM) as a tool,

supporting an address management *practice* is as different from “just DNS” as corporate finance is from bookkeeping.

Our research shows that data center success derives from combining the control of resources managed in-house with the scalability and resilience offered by an outsourced solution. The same benefits can accrue to network services such as DNS and IPAM.

To achieve a high-level of availability and resilience to network outages, companies should consider building network services in an architecture that combine in-house and managed services. A hybrid architecture like this will combine on-premise specialized appliances with off-site managed appliances. This approach allows for separation of the management of internally facing resources (e.g. back-end servers) from externally facing resources (e.g. public-facing servers). It also ensures redundancy so that a company can maintain public presence through the external managed service even if there are network outages or data center outages affecting the internal appliances.

DNS Architecture	Characteristics
In-house	Complete control over domain information, security and compliance. Limited points-of-presence and limited global distribution
Hosted/Managed	Multiple points-of presence, highly available, global load-balancing and re-direction
Hybrid	Best of both worlds: Private DNS space managed internally for full control, public DNS distributed widely for maximum availability and visibility

While a combination of in-house appliances and externally hosted services is a pre-requisite for reliable network services, it is not sufficient. To achieve higher availability and stability, companies need to remove the most common source of problems: manual changes to network services. Inconsistent configuration of network services is commonly identified as a cause for network instability. By automating the configuration of network services, companies can ensure greater consistency of configuration, faster and more predictable upgrades, and more reliable maintenance and change management.

Automation does not mean removing the human element from network management. It means providing the right tools to help people apply policies in a consistent manner. Good automation considers not just the technology but also the people and the process, providing a framework for smooth operations. Automation is how staff scale.

Network Services Are Strategic

Why all the fuss about network services? Because we believe that they are a vital strategic component of a business. A well-managed, resilient and flexible network services infrastructure will enable many of the most important technologies that are of strategic importance to businesses. Conversely, poorly architected and managed network services will delay the implementation of enabling technologies like virtualization, thwart mobility strategies and slow down business innovation.

Virtualization and cloud computing have transformed the face of the data center in the past few years. While virtualization has become ubiquitous, deployed in more than 97% of companies, it is not “fully” deployed. Many companies introduce virtualization and then stumble as they try to virtualize critical enterprise applications. Weaknesses in network services are magnified by the introduction of virtualization, because of the explosion of complexity that occurs in the layer-2 network and the layer-2 to layer-3 boundaries. Data center networks architected for static and inflexible three-tier applications become overwhelmed with the introduction of virtualization and start exhibiting strange instabilities and scalability problems. This is not a new phenomenon: corporate networks were similarly stressed by the introduction of VOIP in the early stages, eight or so years ago. Companies quickly realized that network services infrastructure could either be an enabler to these new technologies, or a serious problem.

If network services are properly architected and managed with consistent processes they can become a strategic enabler for virtualization. Virtualization is not only easier to deploy, it is also easier to monitor and manage if it is based on a solid network services.

Similarly, a good network services infrastructure can be critical to globalization and mobility strategies in large enterprises. It allows companies to manage the large and complex address space that is required to make corporate applications available to mobile users spread around the globe. With a well-architected network services infrastructure, companies can deliver applications with high availability across multiple networks and carriers and to multiple end-user devices.

Well-managed network services also accelerate business innovation, growth, and agility. They allow an organization to rapidly deploy applications without worrying about the availability of IP addresses or the capacity of the IP-address management system. IT managers can use IPAM systems to proactively plan IP address registration and allocation, in advance of demand for IP addresses and with full control and visibility into the current use of the IP address space. Thus, when the IT manager is asked to deploy dozens of servers for a new application, there will be no need for emergency network re-architecture or re-partitioning. There will also be no nasty surprises of over-allocated IP space leading to IP address conflicts, exhausted DHCP servers and seemingly random application outages.

Conclusions and Recommendations

Companies are struggling to handle a huge increase in network complexity and addresses driven by the adoption of new technologies such as VOIP, video, virtualization and RFID. Meanwhile, business trends are spreading the modern enterprise across geographies, with mobility, teleworking, telepresence and virtual collaboration changing the face of business. As these challenges mount, companies must revisit an ad-hoc approach to network services such as DHCP and DNS and review their IP-address management solutions. Network managers must take a holistic view of network infrastructure services, control and automation to protect against instability and foster growth and business agility.

About Nemertes Research: Nemertes Research is a research-advisory firm that specializes in analyzing and quantifying the business value of emerging technologies. You can learn more about Nemertes Research at our Website, www.nemertes.com, or contact us directly at research@nemertes.com.