Network Address Translation FAQ

Contents

Network Address Translation

The Internet is expanding at an exponential rate. As the amount of information and resources increases, it is becoming a requirement for even the smallest businesses and homes to connect to the Internet. Network Address Translation (NAT) is a method of connecting multiple computers to the Internet (or any other IP network) using one IP address. This allows home users and small businesses to connect their network to the Internet cheaply and efficiently.

The impetus towards increasing use of NAT comes from a number of factors:

* A world shortage of IP addresses
* Security needs
* Ease and flexibility of network administration

IP Addresses

In an IP network, each computer is allocated a unique IP address. In the current version of IP protocol, IP version 4, an IP address is 4 bytes. The addresses are usually written as x1.x2.x3.x4, with x1, x2, x3 and x4 each describing one byte of the address. For example, address 16843009 (hex 1010101) is written as 1.1.1.1, since each byte of this address has a value of 1.

Since an address is 4 bytes, the total number of available addresses is 2 to the power of 32 = 4,294,967,296. This represents the TOTAL theoretical number of computers that can be directly connected to the Internet. In practice, the real limit is much smaller for several reasons.

Each physical network has to have a unique Network Number, comprising some of the bits of the IP address. The rest of the bits are used as a Host Number to uniquely identify each computer on that network. The number of unique Network Numbers that can be assigned in the Internet is therefore much smaller than 4 billion, and it is very unlikely that all of the possible Host Numbers in each Network Number are fully assigned.

An address is divided into two parts: a network number and a host number. The idea is that all computers on one physical network will have the same network number - a bit like the street name, the rest of the address defines an individual computer - a bit like house numbers within a street. The size of the network and host parts depends on the class of the address, and is determined by address' network mask. The network mask is a binary mask with 1s in the network part of the address, and 0 in the host part.

To allow for a range from big networks, with a lot of computers, to small networks, with a few hosts, the IP address space is divided into 4 classes, called class A, B, C and D. The first byte of the address determines which class an address belongs to:

* Network addresses with first byte between 1 and 126 are class A, and can have about 17 million hosts each.
* Network addresses with first byte between 128 and 191 are class B, and can have about 65000 hosts each.
* Network addresses with first byte between 192 and 223 are class C, and can have 256 hosts.
* All other networks are class D, used for special functions or class E which is reserved.

Most class A and B addresses have already been allocated, leaving only class C available. This means that total number of available addresses on the Internet is 2,147,483,774. Each major world region has an authority which is given a share of the addresses and is responsible for allocating them to Internet Service Providers (ISPs) and other large customers. Because of routing requirements, a whole class C network (256 addresses) has to be assigned to a client at a time; the clients (e.g.. ISPs) are then responsible for distributing these addresses to their customers.

While the number of available addresses seems large, the Internet is growing at such a pace that it will soon be exhausted. While the next generation IP protocol, IP version 6, allows for larger addresses, it will take years before the existing network infrastructure migrates to the new protocol.

Because IP addresses are a scarce resource, most Internet Service Providers (ISPs) will only allocate one address to a single customer. In majority of cases this address is assigned dynamically, so every time a client connects to the ISP a different address will be provided. Big companies can buy more addresses, but for small businesses and home users the cost of doing so is prohibitive. Because such users are given only one IP address, they can have only one computer connected to the Internet at one time. With an NAT gateway running on this single computer, it is possible to share that single address between multiple local computers and connect them all at the same time. The outside world is unaware of this division and thinks that only one computer is connected.

Security Considerations

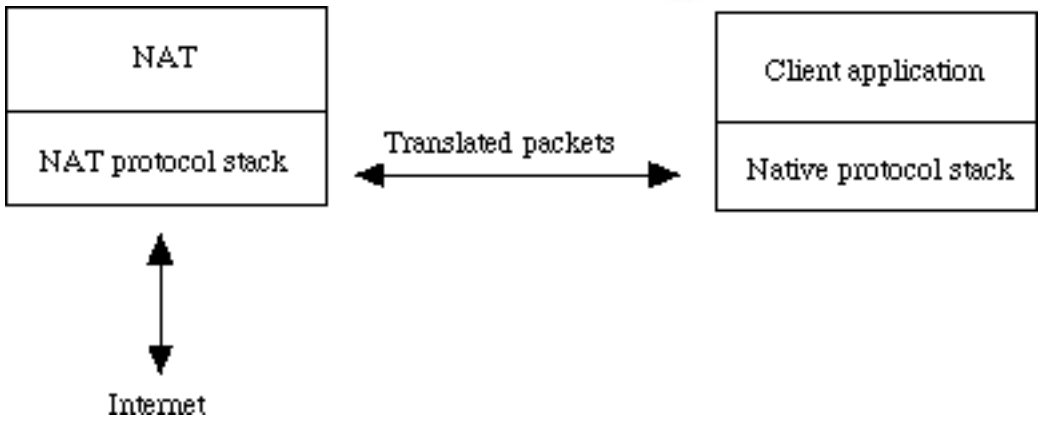
Many people view the Internet as a "one-way street"; they forget that while their computer is connected to the Internet, the Internet is also connected to their computer. That means that anybody with Net access can potentially access resources on their computers (such as files, email, company network etc). Most personal computer operating systems are not designed with security in mind, leaving them wide open to attacks from the Net. To make matters worse, many new software technologies such as Java or Active X have actually reduced security since it is now possible for a Java applet or Active X control to take control of a computer it is running on. Many times it is not even possible to detect that such applets are running; it is only necessary to go to a Web site and the browser will automatically load and run any applets specified on that page.

The security implications of this are very serious. For home users, this means that sensitive personal information, such as emails, correspondence or financial details (such as credit card or cheque numbers) can be stolen. For business users the consequences can be disastrous; should confidential company information such as product plans or marketing strategies be stolen, this can lead to major financial losses or even cause the company to fold.

To combat the security problem, a number of firewall products are available. They are placed between the user and the Internet and verify all traffic before allowing it to pass through. This means, for example, that no unauthorised user would be allowed to access the company's file or email server. The problem with firewall solutions is that they are expensive and difficult to set up and maintain, putting them out of reach for home and small business users.

NAT automatically provides firewall-style protection without any special set-up. That is because it only allows connections that are originated on the inside network. This means, for example, that an internal client can connect to an outside FTP server, but an outside client will not be able to connect to an internal FTP server because it would have to originate the connection, and NAT will not allow that. It is still possible to make some internal servers available to the outside world via inbound mapping, which maps certain well know TCP ports (e.g.. 21 for FTP) to specific internal addresses, thus making services such as FTP or Web available in a controlled way.

Many TCP/IP stacks are susceptible to low-level protocol attacks such as the recently-publicised "SYN flood" or "Ping of Death". These attacks do not compromise the security of the computer, but can cause the servers to crash, resulting in potentially damaging "denials of service". Such attacks can cause abnormal network events that can be used as a precursor or cloak for further security breaches. NATs that do not use the host machine protocol stack but supply their own can provide protection from such attacks:

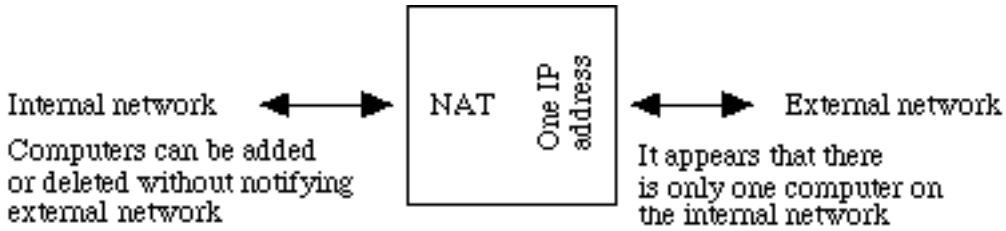


Administrative Considerations

IP networks are more difficult to set up than local desktop LANs; each computer requires an IP address, a subnet mask, DNS address, domain name, and a default router. This information has to be entered on every computer on the network; if only one piece of information is wrong, the network connection will not function and there is usually no indication of what is wrong. In bigger networks the task of co-ordinating the distribution of addresses and dividing the network into subnets is so complicated that it requires a dedicated network administrator.

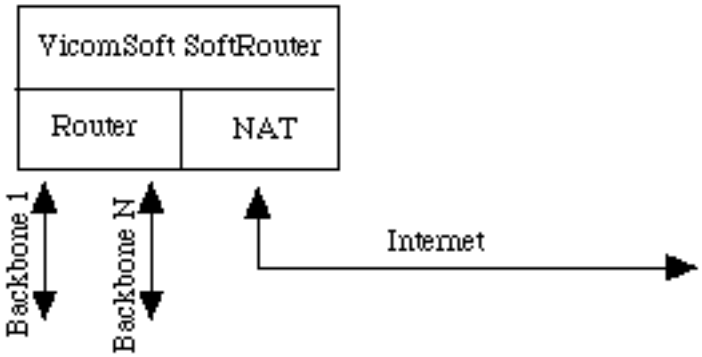
NAT can help network administration in several ways:

* It can divide a large network into several smaller ones. The smaller parts expose only one IP address to the outside, which means that computers can be added or removed, or their addresses changed, without impacting external networks. With inbound mapping, it is even possible to move services (such as Web servers) to a different computer without having to do any changes on external clients.



* Some modern NAT gateways contain a dynamic host configuration protocol (DHCP) server. DHCP allows client computers to be configured automatically; when a computer is switched on, it searches for a DHCP server and obtains TCP/IP setup information. Changes to network

configuration are done centrally at the server and affect all the clients; the administrator does not need to apply the change to every computer in the network. For example, if the DNS server address changes, all clients will automatically start using the new address the next time they contact the DHCP server.
* Many NAT gateways provide for a way to restrict access to the Internet. For example, Vicomsoft InterGate has built-in CyberPatrol filtering, which allows administrators to prohibit access to dubious material.
* Another useful feature is traffic logging; since all the traffic to and from the Internet has to pass through a NAT gateway, it can record all the traffic to a log file. This file can be used to generate various traffic reports, such as traffic breakdown by user, by site, by network connection etc.
* Since NAT gateways operate on IP packet-level, most of them have built-in internetwork routing capability. The internetwork they are serving can be divided into several separate sub networks (either using different backbones or sharing the same backbone) which further simplifies network administration and allows more computers to be connected to the network:



To summarise, a NAT gateway can provide the following benefits:

* Firewall protection for the internal network; only servers specifically designated with "inbound mapping" will be accessible from the Internet
* Protocol-level protection
* Automatic client computer configuration control
* Packet level filtering and routing

NAT and Proxies

A proxy is any device that acts on behalf of another. The term is most often used to denote Web proxying. A Web proxy acts as a "half-way" Web server: network clients make requests to the proxy, which then makes requests on their behalf to the appropriate Web server. Proxy technology is often seen as an alternative way to provide shared access to a single Internet connection. The main benefits of Web proxying are:

* Local caching: a proxy can store frequently-accessed pages on its local hard disk; when these pages are requested, it can serve them from its local files instead of having to download the

data from a remote Web server. Proxies that perform caching are often called caching proxy servers.
* Network bandwidth conservation: if more than one client requests the same page, the proxy can make one request only to a remote server and distribute the received data to all waiting clients.

Both these benefits only become apparent in situations where multiple clients are very likely to access the same sites and so share the same data.

Unlike NAT, Web proxying is not a transparent operation: it must be explicitly supported by its clients. Due to early adoption of Web proxying, most browsers, including Internet Explorer and Netscape Communicator, have built-in support for proxies, but this must normally be configured on each client machine, and may be changed by the naive or malicious user.

Web proxying has the following disadvantages:

* Web content is becoming more and more dynamic, with new developments such as streaming video & audio being widely used. Most of the new data formats are not cacheable, eliminating one of the main benefits of proxying.
* Clients have to be explicitly set to use Web proxying; whenever there is a change (e.g. proxy is moved to a new IP address) each and every client has to be set up again.
* A proxy server operates above the TCP level and uses the machine's built-in protocol stack. For each Web request from a client, a TCP connection has to be established between the client and the proxy machine, and another connection between the proxy machine and the remote Web server. This puts lot of strain on the proxy server machine; in fact, since Web pages are becoming more and more complicated the proxy itself may become bottleneck on the network. This contrasts with a NAT which operates on packet level and requires much less processing for each connection.

NAT Operation

The basic purpose of NAT is to multiplex traffic from the internal network and present it to the Internet as if it was coming from a single computer having only one IP address.

The TCP/IP protocols include a multiplexing facility so that any computer can maintain multiple simultaneous connections with a remote computer. It is this multiplexing facility that is the key to single address NAT.

To multiplex several connections to a single destination, client computers label all packets with unique "port numbers". Each IP packet starts with a header containing the source and destination addresses and port numbers:

Source address	Source port	Destination address	Destination port

This combination of numbers completely defines a single TCP/IP connection. The addresses specify the two machines at each end, and the two port numbers ensure that each connection between this pair of machines can be uniquely identified.

Each separate connection is originated from a unique source port number in the client, and all reply packets from the remote server for this connection contain the same number as their destination port, so that the client can relate them back to its correct connection. In this way, for example, it is possible for a web browser to ask a web server for several images at once and to know how to put all the parts of all the responses back together.

A modern NAT gateway must change the Source address on every outgoing packet to be its single public address. It therefore also renumbers the Source Ports to be unique, so that it can keep track of each client connection. The NAT gateway uses a port mapping table to remember how it renumbered the ports for each client's outgoing packets. The port mapping table relates the client's real local IP address and source port plus its translated source port number to a destination address and port. The NAT gateway can therefore reverse the process for returning packets and route them back to the correct clients.

When any remote server responds to an NAT client, incoming packets arriving at the NAT gateway will all have the same Destination address, but the destination Port number will be the unique Source Port number that was assigned by the NAT. The NAT gateway looks in its port mapping table to determine which "real" client address and port number a packet is destined for, and replaces these numbers before passing the packet on to the local client.

This process is completely dynamic. When a packet is received from an internal client, NAT looks for the matching source address and port in the port mapping table. If the entry is not found, a new one is created, and a new mapping port allocated to the client:

* Incoming packet received on non-NAT port
* Look for source address, port in the mapping table
* If found, replace source port with previously allocated mapping port
* If not found, allocate a new mapping port
* Replace source address with NAT address, source port with mapping port

Packets received on the NAT port undergo a reverse translation process:

* Incoming packet received on NAT port
* Look up destination port number in port mapping table