CHAPTER 2

Understanding Denial of Service

A denial-of-service attack is different in goal, form, and effect than most of the attacks that are launched at networks and computers. Most attackers involved in cybercrime seek to break into a system, extract its secrets, or fool it into providing a service that they should not be allowed to use. Attackers commonly try to steal credit card numbers or proprietary information, gain control of machines to install their software or save their data, deface Web pages, or alter important content on victim machines. Frequently, compromised machines are valued by attackers as resources that can be turned to whatever purpose they currently deem important.

In DDoS attacks, breaking into a large number of computers and gaining malicious control of them is just the first step. The attacker then moves on to the DoS attack itself, which has a different goal—to prevent victim machines or networks from offering service to their legitimate users. No data is stolen, nothing is altered on the victim machines, and no unauthorized access occurs. The victim simply stops offering service to normal clients because it is preoccupied with handling the attack traffic. While no unauthorized access to the victim of the DDoS flood occurs, a large number of other hosts have previously been compromised and controlled by the attacker, who uses them as attack weapons. In most cases, this is *unauthorized access,* by the legal definition of that term.

While the denial-of-service effect on the victim may sound relatively benign, especially when one considers that it usually lasts only as long as the attack is active, for many network users it can be devastating. Use of Internet services has become an

important part of our daily lives. The Internet is increasingly being used to conduct business and even to provide some critical services. Following are some examples of the damaging effects of DoS attacks.

- Sites that offer services to users through online orders make money only when users can access those services. For example, a large book-selling site cannot sell books to its customers if they cannot browse the site's Web pages and order products online. A DoS attack on such sites means a severe loss of revenue for as long as the attack lasts. Prolonged or frequent attacks also inflict long-lasting damage to a site's reputation—customers who were unable to access the desired service are likely to take their business to the competition. Sites whose reputations were damaged may have trouble attracting new customers or investor funding in the future.
- Large news sites and search engines are paid by marketers to present their advertisements to the public. The revenue depends on the number of users that view the site's Web page. A DoS attack on such a site means a direct loss of revenue from the marketers, and may have the long-lasting effect of driving the customers to more easily accessible sites. Loss of popularity translates to a direct loss of advertisers' business.
- Some sites offer a critical free service to Internet users. For example, the Internet's Domain Name System (DNS) provides the necessary information to translate human-readable Web addresses (such as www.example.com) into Internet Protocol (IP) addresses (such as `192.0.34.166`). All Web browsers and numerous other applications depend on DNS to be able to fetch information requested by the users. If DNS servers are under a DoS attack and cannot respond due to overload, many sites may become unreachable because their addresses cannot be resolved, even though those sites are online and fully capable of handling traffic. This makes DNS a part of the critical infrastructure, and other equally important pieces of the Internet's infrastructure are also vulnerable.
- Numerous businesses have come to depend on the Internet for critical daily activities. A DoS attack may interrupt an important videoconference meeting or a large customer order. It may prevent a company from sending out an important document for a rapidly approaching deadline or interfere with its bid for a large contract.
- The Internet is increasingly being used to facilitate management of public services, such as water, power, and sewage, and to deliver critical information for important activities, such as weather and traffic reports for docking ships. A DoS attack that

disrupts these critical services will directly affect even people whose activities are not related to computers or the Internet. It may even endanger human lives.

- A vast number of people use the Internet on a daily basis for entertainment or for communicating with friends and family. While a DoS attack that disrupts these activities may not cause them any serious damage, it is certainly an unpleasant experience that they wish to avoid. If such disruptions occur frequently, people are likely to stop using the Internet for these purposes, in favor of more reliable technologies.

2.1 The Ulterior Motive

Why do attackers seek to deny service? This act, very disruptive in nature, is not always an end in and of itself. What could be the ultimate goal then?

Some of the early DoS attacks were largely proofs of concept or simple pranks played by hackers. The ultimate goal was to prove that something could be done, such as taking a large, popular Web site offline. Such a major achievement brings an attacker recognition in the underground community.

Frequently, attackers would also fight each other for supremacy via denial of service. Internet chat channels were and still are a sought-after resource by the attackers. They are used to coordinate multiple attacking machines and to trade code and illegal information with other attackers. The user who created the channel controls the access to it, and is called a *moderator, operator,* or *owner.* The easy way to take over the channel (and along with it all the attack machines that are controlled via this channel) and to dominate all the communications is to perform a DoS attack on its current moderator. When a moderator's machine goes offline, another user can take over the channel. Besides supremacy, attackers also sought revenge through denial of service. A hacker whose machines were knocked offline by DoS would "return the favor" by attacking the perpetrator. People who dared to speak ill of hackers in public have also felt DoS revenge.

Another frequent motive of DoS attacks is self-described as being political. Individuals or groups who disagree with views or actions of a certain organization (an online media site, a corporation, or a government) have been known to launch DoS attacks against computers and networks owned by this organization.

If the target of the attack is a company, a conceivable motive can be a competitor's wish to gain an edge in the market. So far, no attacks have been proved to have this motive. However, there is a major lack of data on perpetrators and motives of DoS

attacks. The vast majority of attacks are not reported, let alone investigated. Of those that do undergo detailed investigation, only a few contain enough evidence to establish the motive. It is thus quite possible that some companies may resort to such illegal means of driving the competition out of the market.

Recently, a number of attacks have appeared as part of extortion attempts [ZDn04]. The attackers threaten an online business with a denial of service, and a payment is requested for "protection." Sites that refuse the payment are being "persuaded" by small-scale attacks.

2.2 Meet the Attackers

Who are the likely perpetrators of DDoS attacks? We have evidence from studies that thousands of attacks occur on a regular basis, yet very few attackers have been caught and prosecuted. This is partly due to the inability of victims to meet the minimum damage limits necessary to prosecute, or because the victim doesn't feel prosecution is worthwhile or fears negative publicity. Another factor is the ease of performing a DoS attack without leaving many traces for investigators to follow. It is impossible to judge the profile of perpetrators from such a small sample of provable crimes. Still, from the lack of sophistication in many attacks, it is safe to assume that a very large percentage seem to be perpetrated by inexperienced hackers, so-called *script kiddies*. These hackers download crude attack tools from the Internet and use them unaltered. While such attacks can still severely cripple the victim, sufficient traces sometimes exist for investigators to be able to understand much about the attacker. Such crude attacks also frequently generate an easily recognizable traffic pattern that can be controlled by simple filters.

Another type of a DoS perpetrator is a sophisticated hacker who uses several means to obscure her identity and create subtle variations in traffic patterns to bypass defenses. While these attacks are less common than the simple ones, they are particularly vicious and hard to handle. Sophisticated hackers may act on their own accord (when attacking for supremacy in their peer circle or for revenge) or may be hired by an underground movement or a criminal organization.

The most dangerous potential attacker is the nation-state actor that has significant resources and skill available to write his own tools, using sophisticated command and control techniques, and taking advantage of intelligence resources that are hard to come by. Such an attacker could create very subtle effects that are difficult to even notice using common methods or tools. Besides, the monitoring tools may potentially have vulnerabilities themselves that can be exploited to hide the presence of the attack. To

date, no DDoS attacks can be confidently ascribed to such nation-state actors, but they are inherently better at covering their tracks. If no such attacks have occurred yet, they may well occur in the future.

2.3 Behind the Scenes

How do DoS attacks work? As mentioned in Chapter 1, there are two main approaches to denying a service: exploiting a vulnerability present on the target or sending a vast number of seemingly legitimate messages. The first kind of an attack is usually called a *vulnerability attack*, while the second is called a *flooding attack*.

Vulnerability attacks work by sending a few specifically crafted messages to the target application that possesses a vulnerability. This vulnerability is usually a software bug in the implementation or a bug in a default configuration of a given service. Malicious messages by the attacker represent an unexpected input that the application programmer did not foresee. The messages cause the target application to go into an infinite loop; to severely slow down, crash, freeze, or reboot a machine; or to consume a vast amount of memory and deny service to legitimate users. This process is called *exploiting a vulnerability,* and the malicious messages are called the *exploit*. In some cases, vulnerabilities of this kind can be exploited in the operating system, a common piece of middleware, or in a network protocol, as well as in application programs.[1] While it is impossible to detect all vulnerabilities, it can also be quite hard to find new exploits. This means that each vulnerability that is detected and patched is a large gain and a sure step ahead for the defenders.

Flooding attacks work by sending a vast number of messages whose processing consumes some key resource at the target. For instance, complex messages may require lengthy processing that takes up CPU cycles, large messages take up bandwidth, and messages that initiate communication with new clients take up memory. Once the key resource is tied up by the attack, legitimate users cannot receive service. The crucial

[1] For example, some implementations of the 802.11 wireless access protocol have a vulnerability that allows an attack to deny service selectively to one user in the wireless network or promiscuously to all of them. In effect, the attacker can send a packet to the wireless access point that claims to be from another user and that indicates that the user is finished and essentially wants to "hang up" [BS03]. The wireless access point then no longer recognizes communications from the targeted user. That user can reestablish communications with the access point, but the attacker can shut it down again in the same way.

feature of flooding attacks is that their strength lies in the volume, rather than in content. This has two major implications:

1. The attackers can send a variety of packets. The attack traffic can be made arbitrarily similar to the legitimate traffic, which greatly hinders defense.
2. The flow of traffic must be so large as to consume the target's resources. The attacker usually has to engage more than one machine to send out the attack traffic. Flooding attacks are therefore commonly DDoS attacks.

The simplest form of a DDoS attack is merely to send a very large quantity of messages, divided into packets, to a service on the victim machine. Unless something between the attacking machines and the victim drops those request packets, the victim will spend resources attempting to receive and properly handle them. If there are enough of these packets, all of the machine's resources will be spent trying to handle packets that have no value.

Another DDoS option is to attack the victim's network interface. If the network card in the victim's machine can handle only 10 Mbps of traffic, then an attacker needs to merely generate 10 or more Mbps of any deliverable IP packets and send them to the victim. Again assuming that no other entity drops those packets before they reach the victim's interface, they will easily exhaust its network resources and also create a sizable congestion on the path to the victim. If there are a few legitimate packets in addition to the large flood of attack packets, they are unlikely to receive service.

The attacker can also target the local network that attaches the victim to the Internet. If the attacker knows that the victim is attached to a 1-Gbps network segment, then she can send enough packets to the victim or other nodes on the segment to overwhelm it. Most networks become unusable as the traffic offered to them approaches their rated capacity, so little or no legitimate traffic will get through to the victim. In this form of DDoS attack, all of the other nodes on the network segment will similarly suffer. This example illustrates a curious property of DDoS: The damage is inflicted not only on the victim, but also on its legitimate users (who cannot get the service) and anyone else who shares the critical resource. For instance, the attacker may target a network that has the same ISP as you. If the amount of attack traffic is sufficiently high, your services may also be denied.

The above attacks are all based on large volumes of traffic. The attacker can sometimes perpetrate an effective flooding attack with much smaller volumes. If the victim has some service running that requires more time to process a remote request than it takes to generate that request, or that ties up a scarce resource on the server, the

attacker can make use of this asymmetry. Even short or infrequent bursts of malicious traffic will effectively tie up the critical resource. A common example is the TCP SYN flood attack, described in detail in Chapter 4. The attacker floods the victim with TCP SYN packets, which are usually used to initiate new communication. The victim reserves some memory in a limited-size buffer for each new communication request, while the attacker can send out those requests without any memory cost. This asymmetry helps the attacker disable any new communication during the attack, while sending very few TCP SYN packets.

This discussion illustrates the fact that the line between vulnerability and flooding attacks is thin, and many attacks may well fall into both the vulnerability and flooding categories.

2.3.1 Recruiting and Controlling Attacking Machines

DDoS attacks require engagement of multiple machines, which will be sending the attack traffic to the victim. Those machines do not belong to the attacker. They are usually poorly secured systems at universities, companies, and homes—even at government institutions. The attacker breaks into them, takes full control, and misuses them for the attack. Therefore, the attacking machines are frequently called *zombies*, *daemons*, *slaves*, or *agents*. In this book we use the term *agents*.

How does the attacker gain control over machines that belong to others? The agents are usually poorly secured machines—they do not have recent patches and software updates, they are not protected by a firewall or other security devices, or their users have easily guessed passwords. The attacker takes advantage of these well-known holes to break in. Unpatched and old software has well-known vulnerabilities with already-written exploits. These belong to a specific kind of vulnerabilities—once exploited, they allow the attacker unlimited access to the system, as if he had an administrator's account. Accounts with easily guessed passwords, such as combinations of users' names or dictionary words, allow another easy way into the machine. There are several password-guessing tools that will quickly reveal if any of the accounts on a system have weak passwords.[2] For example, *Phatbot* will attempt to connect and log in to Windows hosts using a set of several dozen very commonly chosen passwords. Even if these programs find accounts that do not have administrator privileges, this access can still be misused

[2] A weak password is one that is easily guessed, by either a human guesser or an automated program that tries many possible passwords, such as all very short passwords, all the words in the dictionary, or the most commonly used men's and women's names.

for a DDoS attack or, by exploiting other vulnerabilities, can elevate privileges to the administrator level.

Once the attacker has gained control of the host, she installs the DDoS attack agent and makes sure that all traces of the intrusion are well hidden and that the code runs even after the machine is rebooted.

DDoS attacks frequently involve hundreds or thousands of agents. It would be tedious and time consuming if the attacker had to manually break into each of them. Instead, there are automated tools that discover potential agent machines, break into them, and install the attack code upon a single command from an attacker, and report success back to her. Such tools can easily be downloaded from the Web or acquired from Internet chat channels. In addition to recruiting the collection of agents, automated tools also facilitate the control of this network by keeping track of the agents and providing easy ways of delivering commands to all of them at once. The attacker needs only to issue a single command and have all agents start the flood to the given victim.

2.3.2 Hiding

The attacker further hides her identity by deploying several layers of indirection between her machine and the agents. She uses one or several machines that deliver her commands to the agents. These machines are called *handlers* or *masters*. In this book we use the term *handlers*. Figure 2.1 illustrates this handler/agent architecture.

Another layer of indirection consists of the attacker's logging on to several machines *in sequence*, before accessing the handlers. These intermediary machines between the attacker's machine and the handlers are called the *stepping stones*, and are illustrated in Figure 2.2.

Both handlers and stepping stones are used to hinder investigation attempts. If authorities located and examined an agent machine, all its communication would point to one of the handlers. Further examination of the handler would point to a stepping stone, and from there to another stepping stone. If stepping stones are selected from different countries and continents (and they usually are), it becomes very difficult to follow the trail back to the attacker's machine and unveil her identity.

Another means of obscuring the attack is through the use of *IP spoofing*. Each packet in the Internet carries some control information preceding the data—an *IP header*. One field in the IP header specifies the address of the sender—the *source IP* field. This information is filled in by the machine that sends the packet (an action similar to putting a return address on a letter), and is used by the destination, or the routers on the path to the destination, to send replies back to the source. Attackers

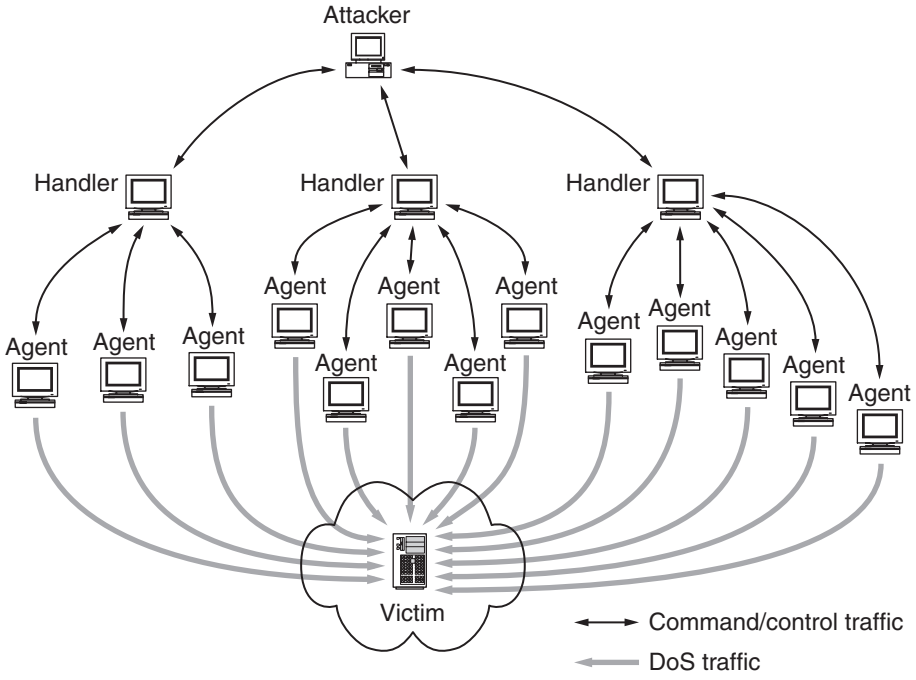

Figure 2.1 Handler/agent architecture

commonly forge this field to achieve impunity for the attacks and hinder the discovery of agent machines. IP spoofing also greatly complicates some DDoS defense approaches that rely on a source address for differentiation between legitimate clients and attackers. With IP spoofing, an attacker easily assumes the identity of a legitimate client or even several of them.

2.3.3 Misusing Legitimate Services

IP spoofing creates an opportunity for fooling noncompromised and otherwise perfectly secure machines into participating in a DDoS attack. The attacker chooses a publicly available service or protocol, such as the Domain Name System (DNS), Web, or ping, and sends service requests to many such servers, forging the source address of the victim. Servers then reply back to the victim, and this flood of replies creates denial of service. This type of attack is called a *reflection attack*, and the servers participating in it are called *reflectors*. Of special interest to the attacker are services that can generate lengthy

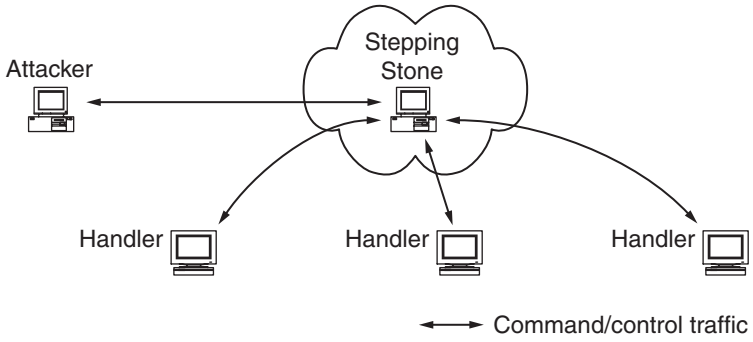

Figure 2.2 Illustration of a site hosting a stepping stone

or numerous replies for a few short requests. This is called *amplification*, and it enables the attacker to create a DoS effect with a few small packets.

2.4 Distribution Effects

Denial of service is possible without using distributed techniques, but it poses a challenge for an attacker. For example, imagine that a DoS attack based on pure flooding originates at a single machine with a 10-Mbps link and is directed toward a victim machine that has a 100-Mbps link. In an attempt to overwhelm the victim's link, the attacker will flood his own network and deny service to himself. To successfully disrupt the victim's communication, the attacker must compromise an agent machine that has more network resources than the victim. Locating and breaking into such a machine may prove difficult, especially if the target of the attack is a well-provisioned site.

However, consider what happens if the same attack is performed in a distributed manner, say, by a hundred machines. Each machine now sends 1 Mbps toward the victim. Assuming all hundred machines have 10-Mbps links, none of them generates enough traffic to cause serious harm to its own local network. But the Internet delivers all attack traffic to the victim, overwhelming its link. Thus, the victim's service is denied, while the attackers are still fully operational.

Distribution brings a number of benefits to the attacker:

- A typical server machine has more computing, memory, and bandwidth resources than a typical client machine. An attacker with control of only one client machine would thus have difficulty overwhelming the server's resources without first overwhelming his own. By using distributed techniques, the attacker can multiply the

resources on the attacking end, allowing him to deny service to more powerful machines at the target end.

- To stop a simple DoS attack from a single agent, a defender needs to identify that agent and take some action that prevents it from sending such a large volume of traffic. In many cases, the attack from a machine can be stopped only if the machine's human administrator, or network operator, takes action. If there are 10, 100, or 1,000 agents participating in the attack, however, stopping any single one of them may provide little benefit to the victim. Only by stopping most or all of them can the DoS effect be alleviated. Getting thousands of people to take some action to stop an attack from their machine is an overwhelming challenge.[3]
- If the attacker chooses agents that are spread widely throughout the Internet, attempts to stop the attack are more difficult, since the only point at which all of the attack traffic merges is close to the victim. This point is called the *aggregation point*. Other nodes in the network might experience no telltale signs of the attack and might have difficulty distinguishing the attack traffic from legitimate traffic. Thus, they cannot help defend against the attack.
- In a DoS attack executed from a single agent, the victim might be able to recover by obtaining more resources. For example, an overwhelmed Web server might be able to recruit other local servers to help handle the extra load. Regardless of how powerful a single agent might be, the defender can add more capacity until he outstrips the attacker's ability to generate load. This approach is less effective in defending against DDoS attacks. If the defender doubles his resources to handle twice as many requests, the attacker merely needs to double the number of agents—often an easy task.

Another aspect makes both DoS and DDoS attacks hard to handle: Defenses that work well against many other kinds of attacks are not necessarily effective against denial of service. For years, system administrators have been advised to install a firewall and keep its configuration up to date, to close unnecessary ports on all machines, to stay current with patches of operating systems and other important software, and to run intrusion detection systems to discover any attacks that have managed to penetrate the outer bastions of defense. Unfortunately, these security measures often will not

[3] Alternatively, defenders might attempt to locate a handler machine and command agents to stop flooding. This is a challenging task, too, since the attacker may have multiple handlers or use a legitimate service (e.g., IRC) instead of a handler, and the agent commands may be encrypted or password-protected.

help against denial of service. The attack can consist of traffic that the firewall finds acceptable, probably because it bears a close resemblance to legitimate traffic. Since the DoS attack merely needs to exhaust resources, it can work on any port left open, including those that must be open for a node to do its normal business. Attackers can perform DoS attacks on machines that have no vulnerabilities (by the standard definition of that term), so patches to close vulnerabilities may not help. Also, intrusion detection systems are of limited value in dealing with DoS, since, unlike break-ins and thefts, DoS attacks rarely hide themselves. After all, their whole purpose is to interrupt normal business, an event that will usually be noticed.

2.5 DDoS: Hype or Reality?

The issues described in the previous section make DDoS attacks a frightening possibility. Yet researchers in computer and network security are aware of many frightening possibilities that never come to pass. Are security researchers merely alarming the public with claims of the dangers of DDoS?

Unfortunately, DDoS attacks are not speculation or fiction. They occur on a daily basis, directed against a wide range of sites. Chapter 3 details, in timeline fashion, a large number of representative attacks. The details of these attacks will be left to that chapter, while some specifics will be mentioned in this chapter. In addition to several well-known occurrences of DDoS attacks that were widely reported in the press, there are scientific studies of the frequency of these attacks that demonstrate the reality of the problem (see Appendix C for a summary of these studies).

2.5.1 How Common Are DDoS Attacks?

There are some forms of cyberattacks that receive a lot of publicity because they generate a few high-profile incidents, even though these types of attacks do not actually occur that often. Unless these incidents are particularly disastrous, the overall impact of the attacks is more related to publicity than large amounts of damage done to many businesses or individuals.

DDoS attacks do not fit that category. A number of recent studies have demonstrated that DDoS attacks are extremely common in today's networks. Given that they are usually quite effective and perpetrators are rarely caught, there is reason to believe they will become even more popular in the future.

Measuring the frequency of any form of attack in the Internet is difficult. Victims do not always realize that they are under attack. Even if they do, they often fail to

report the attack to any authority. A number of organizations use survey techniques to gain some insight into the prevalence of different kinds of cyberattacks and the amount of damage they do. One example is the FBI's annual report on cybercrime, based on information provided by nearly 500 organizations. In the 2004 report, nearly a fifth of the respondents who suffered financial loss from an attack had experienced a DoS attack. The total reported costs of DoS attacks to these companies was over $26 million. Denial of service was the top source of financial loss due to cybercrime! These surveys are often criticized because their methodology is unavoidably subject to certain limitations, but relatively little better data exists.

The methods used in these surveys do not differentiate between distributed and nondistributed DoS attacks, since the technology for making the distinction is in its infancy. In the meantime, researchers have used a variety of techniques to estimate data on the frequency of DDoS attacks and their other characteristics.

For example, Farnam Jahanian of the University of Michigan has been able to observe network activities in the MichNet ISP. This network provider offers ISP service to government and nonprofit organizations in the state of Michigan, including most educational institutions in that state. Over the course of time, Jahanian's team has gathered data suggesting that DDoS attacks are quite common and are increasingly sophisticated. Jahanian's full results have not yet been published; however, a presentation covering some of his results can be found at `http://www.arbor.net/downloads/nanogSlides4.pdf`.

A number of researchers have investigated various technical means to deduce information about the prevalence and character of DDoS attacks in the Internet [DLD00]. CAIDA (the Cooperative Association for Internet Data Analysis), for example, used a technique called *backscatter*. Full details of this technique and CAIDA's results can be found in Appendix C. Their results suggest that during a three-week observation period in 2001 there were around 4,000 DDoS attacks per week on Internet nodes.

For reasons covered in Appendix C, CAIDA's numbers are certainly an underestimate. Jahanian's results can be interpreted to suggest that the CAIDA figure of 4,000 attacks per week would be more realistically set at 12,000 attacks per week, even leaving aside some classes of DDoS attacks. Further, other data suggests that DDoS attacks have become more common since 2001.

If DDoS attacks are so common, why do we not hear more about them? Evidence gathered by CAIDA and Jahanian suggests that most DDoS attacks are launched against fairly small targets (home machines, for example) for short durations. Some have speculated that many of the incidents represent hackers attacking each other, though

too little evidence exists to come to any strong conclusion on this point. Short durations can cause a DDoS attack to appear to be no more than another network glitch. When a user clicks on a link and receives no response for a minute or two, he is more likely to conclude that the server is busy or that there are general network congestion problems, rather than that he (or, more likely, the server) is suffering a DDoS attack. Thus, in many cases DDoS attacks may pass unnoticed.

If many DDoS attacks are not even noticed, how seriously should we regard the problem? First, there is a significant and growing number of high-profile incidents of serious, persistent, powerful DDoS attacks clearly meant to deny service to important sites. Second, remember that the small, short attacks are typically small and short because that was what the attacker wanted to do, rather than what he could do. A DDoS agent network can continue its attack for hours, or perhaps even indefinitely. And attackers can easily gather huge agent armies. The techniques are already well known and of proven effectiveness. All that remains is a sufficient motive for them to be widely used for destructive purposes.

2.5.2 The Magnitude of DDoS Attacks

Another potentially measurable dimension of a DDoS attack is its size. The size of an attack can be measured in the traffic it generates or in the number of sites participating in the attack. It can also be measured in its duration, a characteristic that some DDoS studies have addressed.

The built-in statistics capabilities of the Shaft attack tool [DLD00] allowed researchers to estimate the magnitude of a given attack in late 1999, at 4.5 Mbps emanating from a single DDoS agent in a network of about 100 agents (see also Figure 4.11 in Chapter 4). Also, MultiRouter Traffic Grapher (MRTG) measurements [Oet] from an actual attack in May 2001 collected close to the target location provide a lower estimate for the inbound attack traffic volume of about 25 Mbps (see Figure 4.13 in Chapter 4). The lower estimate is due to the measurement equipment collapsing intermittently under the heavy load.

DDoS attacks that have taken out large network links in the past, such as an attack on Australian Uecomm, have involved volumes of up to 600,000 pps [Gra]. In attacks on the DNS root servers in 2002, each server received 100,000 to 200,000 pps [Nar]. In some cases, such as the Al-Jazeera attack in 2003, the attackers added attack volume as the defenders added capacity to handle traffic. This shows that attackers can easily increase the attack strength when necessary, so the measured attack magnitudes have more to do with what the attacker feels is required than with the maximum amount

that he can generate. In fact, many attacks may have specifically used a set of moderate-sized discrete attack networks so as to not expose all of them at one time. More recent attackers have learned that it is wasteful to use all of their resources at one time and instead ramp up an attack slowly to maximize how long the attack can be maintained in the face of attrition of the agents.

The backscatter approach used by CAIDA can also estimate the volume of attacks. (Again, for details on how this can be done, see Appendix C.) Taking into account certain limitations of the approach that might lead to underestimates, half of the attacks they observed caused volumes of 350 pps or more. Depending on the target's capabilities, the type of packet, and the target's defenses, this volume is often enough to deny service. The largest volumes CAIDA deduced were hundreds of thousands of packets per second. For example, in the TCP SYN flood attacks against SCO in December 2003, CAIDA estimated that SCO's servers received as many as 50,000 pps at one point and dealt with a total of over 700 million attack packets over a 32-hour period. They estimated this peak rate of 50,000 pps yielded "approximately 20 Mbits/second of Internet traffic in each direction, comparable to half the capacity of a DS3 line (roughly 45 MBits/second.)" [MVS01].

In terms of the number of machines involved in an attack, statistics are harder to come by. It is clear from evidence gathered by the University of Minnesota, which suffered one of the first DDoS attacks in 1999, that DDoS attack networks could be assembled from well over 2,200 systems using only partially automated agent recruitment methods. This minimum number is known because that attack did not use IP spoofing. In attacks in which some form of IP spoofing is used, merely counting the number of IP addresses observed during a particular DDoS attack will grossly overestimate the number of nodes involved.

Another approach is to deduce the number of machines from the observed volume. The largest attack rate observed by CAIDA was estimated to be 679,000 pps. How many packets a machine can generate per second depends on several factors, including its CPU speed and network connectivity. For machines with 10-Mbps links to the Internet, generating 20,000 pps is probably near their maximum capability. So if we assume the largest attack observed by CAIDA was performed by a group of such machines, there had to be at least 30 or 40 of them. For the DNS server attack mentioned above, there had to be at least 90 of them. Many machines have substantially lower speed Internet connections, and if these machines are used as agents, many more of them would be required to achieve these rates. For example, if all agents used 56 Kbps links to connect to the Internet, CAIDA's largest observed attack would have involved at least 5,800 agents. The actual number of agents used in this attack is probably between

these ballpark figures. Reflected attacks, where attacking hosts send out forged attack packets that are reflected off a very large number of legitimate servers around the world, greatly amplify the attack. One such attack against futuresite.register.com involved a very small number of attacking hosts, but was still able to generate 60 to 90 million bits per second flooding the victim.

One might wonder where the DDoS agents come from. Most experts believe that very few attackers use their own machines to launch DDoS attacks, since doing so would increase their risk of being caught. Instead, they compromise other machines remotely and use them to launch the attack. If compromising a remote machine were a difficult process requiring extensive human intelligence and attention, this factor would limit the seriousness of the DDoS threat. However, experience has shown that automated techniques are highly effective at compromising remote sites, which can then be used to launch DDoS attacks.

Just to give an idea of how easy it is to compromise a large number of hosts, here are some figures:

- Microsoft announced that their MSBlast cleanup tool was downloaded and used to successfully clean up 9.5 million hosts from August 2003 to April 2004, an average of approximately 1 million compromised computers per month (see `http://zdnet.com.com/2100-1105-5201807.html?tag=nl`).
- Microsoft announced in May 2004 that they had cleaned up 2 million *Sasser* infected hosts (see `http://www.securityfocus.com/news/8573`).
- The same news story reports Symantec had identified a bot network of 400,000 hosts.
- A network administrator in the Netherlands has identified between 1 million and 2 million unique IP addresses associated with *Phatbot* infections. *Phatbot* has features to harvest MyDoom- and Bagel-infected hosts, among other infection vectors (see `http://www.ladlass.com/archives/001938.html`).

Probably the most common method of recruiting agents is to run an automated program that scans a large IP address range attempting to find machines that are susceptible to well-known methods of compromise. These programs, called automated infection toolkits, or auto-rooters (after the name of the system administrator account on Unix systems, *root*, also the hacker verb meaning "to compromise or gain elevated privileges on"), are generally quite successful in finding large numbers of vulnerable machines, particularly if they are updated to include newly discovered vulnerabilities that are less likely to have been patched.

The ultimate in automation is an Internet worm—a program that looks for vulnerable machines and infects them with a copy of its code. Worms propagate extremely rapidly. Some worms have used their armies of infected machines specifically to perform DDoS attacks. The worm can even carry the code to perpetrate the DDoS attack. For example, Code Red was designed to perform a DDoS attack from all the nodes it compromised on a particular IP address. Code Red succeeded in infecting over 250,000 machines, by some estimates. Code Red II infected as many as 500,000 machines. Estimates for the number of machines infected by the *W32/Blaster* and *W32/Sobig*.F worms run from the tens of thousands to a few hundreds of thousands, and some reports refer to these numbers as "small." Sasser infected at least 2 million hosts, judging by Microsoft's report (`http://www.securityfocus.com/news/8573`). Thus, it is quite realistic to envision DDoS attacks originating from hundreds of thousands, even millions of points in the Internet.

2.6 How Vulnerable Are You to DDoS?

If you accept that DDoS attacks are a real threat to some Internet sites, the next question likely to come to mind is: How vulnerable is my site? The simple answer is that if your site is connected to the Internet, you are a potential target of a DDoS attack. A DDoS attack can target any IP address and, if the attack is strong enough, it is likely to be successful. Large and small businesses, ISPs, government organizations that rely on networking, and even private individuals are among those who may be damaged by a DDoS attack. The more use you have for the Internet in your enterprise, the greater the damage you will suffer if a DDoS attack takes it offline for an extended period.

Even if your machine sits behind a NAT box,[4] a firewall, or some other form of protection that prevents arbitrary traffic from being directly routed to it, you may still be vulnerable to the more sophisticated DDoS attacks. A sophisticated attacker can replay or spoof traffic that should go to your node or indirectly subject you to denial of service by overloading the NAT box, firewall, router, or network link.

[4] A Network Address Translation (NAT) box is a firewall-like host acting as a gateway to a network. All packets leaving the network pass through the NAT box and have their source addresses replaced by the address of this box. A reverse transformation is applied to destination addresses of incoming packets—the address of the NAT box is replaced with the appropriate address of a machine inside the network. The NAT technique enables a network to hide its internal structure—the only address that external users ever see is that of the NAT box.

Further, as we previously discussed, careful system and network administration will not necessarily save you from an attack. While some fixes will prevent vulnerability attacks, your site will still be susceptible to large flooding attacks.

Heavy provisioning, in the form of ample server and network capacity, can protect you from many flooding DDoS attacks, but cannot guarantee your immunity. Any realistic amount of capacity you provide can be overcome if an attacker recruits enough machines to press his attack against you. Reflect on how heavily you would have to provision yourself to withstand a DDoS attack by the million-plus *Phatbot* network reported earlier.

Nonetheless, there are things you can do to decrease your vulnerability to DDoS attacks and make you a less attractive target. Heavy provisioning helps, since it rules out casual attacks by hackers who have only one or two dozen agent machines at their disposal. Closing vulnerabilities also helps, since it fends off vulnerability attacks. If keeping a low profile on the network is an option for your organization, doing so requires the attacker to find some obscure information before he can launch his attack. There are practical steps to take to strengthen your network and also efficient attack responses that alleviate the DoS effect. We will discuss these in more detail in Chapter 6. Chapter 7 looks at research approaches that may lead to new DDoS defense tools in the future. A number of commercial products have successfully defended against many forms of DDoS attack; we will discuss some of them in Appendix B.

Generally, the evidence suggests that practically all DDoS attacks that occur are not nearly as bad as catastrophic worst-case-scenario thinking suggests they could be. Even some of the high-profile attacks on major Internet sites were not that difficult to handle once the defenders were aware of the nature of the attack and had a little time to respond to it. If you depend on continual Internet availability of your resources, you are almost certainly in danger from DDoS attacks; but with a little knowledge, forethought, and vigilance you can prevent DDoS attacks on your site from becoming disasters.

Even if you are not particularly dismayed by the prospect of being a DDoS victim, another element of DDoS attacks might cause you trouble. To perpetrate a strong DDoS attack, the attacker typically compromises a large number of machines. If your machine is among them, at best you are unwillingly sharing your resources with a criminal who definitely doesn't have your best interests at heart. At worst, you may find yourself partially liable for some of the damages done by his attack, or your vital data may be stolen or damaged by the attacker who has taken over your machine. The value attackers obtain by performing DDoS attacks on others has made such criminals more motivated to compromise ever larger armies of agent machines, meaning that your machine has become more likely to be taken over by an outside party.