Counter-Attacks for Cybersecurity Threats

Andrew Hoskins (UW)

Yi-Kai Liu (UCSD)

Anil Relkuntwar (UW)

Final Project for

Cyber Security and Homeland Security (CSE P 590TU)

Dec. 7, 2005

Approximate division of labor:

- I. Introduction – Yi-Kai Liu
- II. History – Andrew Hoskins
- IV. Technical aspects
    - Cybersecurity Incident Response – Anil Relkuntwar
    - Anti-worms – Andrew Hoskins, Anil Relkuntwar
    - Botnet, spam – Anil Relkuntwar
    - Misleading hackers, tracking files, fake vulnerabilities – Andrew Hoskins
- V. Feasibility
    - General observations – Yi-Kai Liu
    - Anti-worms – Andrew Hoskins, Anil Relkuntwar
    - Spam, botnet – Anil Relkuntwar
    - Misleading hackers, tracking files, fake vulnerabilities – Andrew Hoskins
- VI. Legal issues
    - Computer crimes, self defense – Yi-Kai Liu
    - Other considerations, specific scenarios – Andrew Hoskins, Yi-Kai Liu
- VII. Recommendations for the future
    - Improving cooperation – Yi-Kai Liu
    - Attribution, infrastructure for counter-attacks – Anil Relkuntwar
- VIII. Conclusions – Andrew Hoskins

I. Introduction

Many people have proposed methods of "counter-attack" or "active response" to deal with today's cybersecurity threats. There is a common belief that traditional approaches to the problem are inadequate. Patching software vulnerabilities is labor intensive, so many people don't do it until there's an attack. The legal system is slow to respond, courts lack the necessary technical expertise, and prosecutions do not occur often enough to deter attacks.

At the same time, the Internet has created new threats, such as distributed denial-of-service (DDOS) attacks, worms, and botnets. Attackers have obviously learned to exploit the Internet's global connectivity and pervasive insecurity. Counter-attacks offer a way to take back those technical advantages for the purpose of improving security. At the same time, legal and ethical questions remain. When is a counter-attack justified as self-defense, and when is it vigilante justice?

This paper examines these issues, in the context of commercial (i.e., civilian) cyber-security. We do not discuss information warfare, or cyber-attacks to enforce other laws (such as copyrights or software licenses). Section II gives a brief history of cyber counter-attacks. Section III describes several specific attacks, and sections IV through VI analyze the technical details, practical feasibility and legal issues of these attacks. Section VII presents some recommendations for the future, and section VIII concludes.

II. History

The history of successful counterattacks in the world of cybersecurity is a short one. There are an increasing number of papers and publications which make proposals, but concerns over legal issues as well as technical challenges have kept the list of successes small.

In Sept. 1998, the Pentagon is reported to have counterattacked some activists who were doing a Denial of Service attack of the Department of Defense's websites, responding to their requests with an applet which ran on the attackers' machines and forced them to reboot.[1]

In 1999, a California ISP Conxion wrote a script which caused DoS packets sent to the World Trade Organization's website to be sent back to the attackers.[2] Of course, such an attack would not have worked if the attackers had used what's known as "IP address spoofing", which would have made the packets appear that they came from a different computer from where they did.

In 2001, several defenses were created in response to the Code Red II worm. One, called CRclean, if it received a probe from the worm, would make use of a back door which the worm placed on the infected machine to load a neutralizing agent and halt the spread of the worm. But it would also install itself on that machine. Another, called CodeGreen, would actually scan the internet for computers with the IIS vulnerability that allowed CodeRed, then download the patch and place it on the machine, and clean up the back doors.[3] It is clear that there are serious legal and ethical issues with both of these worms, even the former, which is still a worm even though it only reacts to attacks.

In 2003, the Welchia worm was released onto the Internet. Welchia would remove the Blaster worm and patch the vulnerability that Blaster used to spread. However, Welchia also caused severe network congestion.[4]

In late 2004, Lycos Europe got fed up with spammers and launched a popular campaign called "Make Love Not Spam" which attracted over 100,000 users. Users could install a screen saver on their desktop which would send requests to websites which were known to advertise via spam. This effort was sucessful in causing some of the sites to change locations. (http://www.makelovenotspam.com) Lycos reportedly got around the illegality of DDoS attacks by claiming to only take 95% of the target sites' bandwidth, though, David Dittrich finds this reasoning dubious at best. Lycos ended this campaign in December 2004.[5]

Some companies such as Symbiot Security are currently developing counter attack tools and devices, but there is a lot of skepticism about deployment of devices because of liability concerns.[6]

III. Scenarios

Anti-worms: An anti-worm is a tool that counter-attacks a worm-infected host; ideally it disables the worm but leaves the host running. An anti-worm may be a worm itself (i.e., it may self-propagate). It may also be used against other kinds of malware.

Spam: When server receive spam, it retaliates by bouncing e-mails or an email to Email Server's administrator for remedial action. (Even if sender addresses are spoofed, one may be able to identify the last mail server that forwarded a message.)

Botnet: A compromised host is monitored for "Home Callback" thus identifying the attacker. Strike back is done by

1. Taking down the Home machine.

2. Taking control over the command channel and then using it to neutralize other compromised hosts and strike back at the attacker.

Mislead a Hacker's Investigations: A hacker is scoping out a website to deduce its structure and vulnerability. There are various techniques to mislead the hacker or break or mislead the hacker's automated tools.

Tracking File Transfers: To protect a file, one can surreptitiously enclose a beacon which, upon a copy or install, will announce its presence to the owner of the file. This can help with Digital-Rights Managed files or private files which have been leaked.

Fake Vulnerabilities: A honeypot (say, with bees,) can provide the attacker with a malicious command shell that, unless the hacker is extra careful, will compromise the hacker's machine.

Distributed Denial of Service: A user community can be enlisted to strike at a known malicious website, for example, while running screensavers.

IV. Technical aspects of scenarios

Cybersecurity Incident Response

Counter-attacks are part of the overall process of cybersecurity incident response. This consists of five stages:

1. Intrusion detection.
2. Identifying the attacker.
3. Defending against the attack.
4. Designing and deploying a counter attack.
5. Post-conflict investigation and support.

Intrusion Detection: Intrusion detection first identifies if there is an intrusion or intrusion attempt in the network or machine connected to network. Monitoring can be done on the network or on end-hosts. Firewalls and virus scanners also contribute. There are various techniques, such as anomaly detection, and policy-based approaches.

Identifying the Attacker: This very important step towards counter-attack, basically knowing who your enemy is and where is he located. Identifying the attacker can be difficult, since an attacker may use one compromised host to attack another. Also, with today's network architecture it can be quite difficult to locate the real attacker. Packets can be sent with spoofed source addresses, and tracing them through multiple hops and different autonomous systems can be a complicated process. Finding out the specific server and location can be critical in determining what the counter attack would be.

Defending against the attack: Before an attack, software patches are used to fix vulnerabillities. More interesting are defenses once an attack against a real vulnerability has happened. Diversions can be set forth to surprise an attacker. Or once the attack has been detected, countermeasures can be put in place to stop further damage. Also, diversion tactics like Honey Pots can be used to keep the rest of the network safe. The information from these diversions (or information gained during the delays caused by these tactics) can be harnessed to engineer a specific counter attack.

Designing and deploying a counter attack: Counterattacks take a surprising variety of forms. Designing a counter attack should take into account the following factors: the method of attack; the support of the network infrastructure (i.e. are ISPs performing egress filtering); collateral damage; and costs.

Post Counter attack investigation and support: In order to aid the post counter attack investigation a counter attack should have inbuilt mechanism to log its trail and its operation history. Ideally, it will also provide a means of discovering the hacker's techniques and identity. The more risk hackers have of being discovered, the less likely they are to go hacking.

Keeping the above in mind, we present details of each of the scenarios listed above

Anti-worms

A worm spreads itself by sending out packets which look for certain responses from computers. It will analyze those responses to deduce if it thinks the replying machine has the vulnerability that it exploits. These packets represent a worm's 'signature.' The machine the worm is targeting can detect the signature of the worm and thus know its mode of propagation. An anti-worm is generated, which exploits the same vulnerabilities as the original worm or uses the back doors opened by the original worm. The antiworm would either fix the attacking host, neutralize the worm, or bring it down gracefully, potentially informing the owner. It may also install itself for further propagation.[7]

In addition to the stated goal, an antiworm adheres to following guidelines to minimize collateral damage:

1. It is bug-free and does leave other vulnerabilities that could be exploited further.

2. It can be uninstalled gracefully when owner asks for it.

3. It makes minimal changes to host that it operates on.

4. It should not be able to be used for propagating malicious content.

The countermeasure, or 'Neutralizing Agent' used, can take several forms. For example, it could patch the vulnerability and remove the botnet or worm. Less invasive changes include blocking the port over which the worm sends messages, or using a mutex and rebooting the machine to lock the botnet from being able to execute.[8] Also, one would have to remove any backdoors left by the worm.

In addition, deployment strategy of the antiworm will determine its effectiveness and collateral damage it may cause.[7]

1. Active deployment: The antiworm is itself a worm; it propagates across the network by scanning and seeking exploited vulnerabilities. This is the fastest way to remove the original worm from the network, but this also causes the most collateral damage by affecting the network traffic a lot.

2. Passive Deployment: The anti-worm waits on a host till another host attacks it, and then the antiworm cures only the attacking host. Collateral damage is bare minimum, but the anti-worm is left behind on the new host.

3. Controlled Deployment: The anti-worm propagation is control via some statistic of number of infected/vulnerable host to number host it cured. This can be very effective and efficient.

4. IDS base deployment: The anti-worm uses the intrusion detection sensors which monitor traffic and may deploy an anti-worm against an attacker. As with passive deployment, collateral damage is limited to computers actively launching attacks, but this is less invasive because it does not leave a worm, it just neutralizes the machine.

Spam

Spam is unsolicited email. A virus is a program which, when executed, installs malware such as root-kits, key-stroke loggers, etc. These programs are then used to send sensitive information to the attacker without the user's consent. Both spam and viruses spread through email and require human interaction to activate them.

Email servers as well as the hosts can detect spam with the help of spam detectors like Symantec's TurnTide. Once the spam is identified one counter attack is achieved by automatic flooding of the Sender Server with bounced emails and mail notification to server's administrator. (Even if sender addresses are spoofed, one may be able to identify the last mail server that forwarded a message.)

Botnet

An attacker establishes a botnet by first spreading a virus or worm that takes over the vulnerable hosts. Each worm or virus then calls Home to tell that it has successfully captured the Host. It also establishes a command channel such that the Attacker through the home can control the host. Botnets can be used for DDOS attacks, phishing etc.

Counter Attack:

1. Counter attack involves identifying the attacker by monitoring the compromised Host and intercepting its communication to the Home machine. The Home machine is probably the attacker's base or intermediate node for communication between the host and the attacker base machines. Once that machine is identified, you hack back into that machine bring it down. For hacking back, you can use same techniques as used by the hackers to take control of the attacking machine.

2. Taking down the Home machine only removes the attacker for a short duration, the network of compromised hosts still exists. A more effective counter attack would be to takeover or monitor the command channel for communication between the Home and other compromised hosts. If possible, identify other compromised hosts and fix them or take them down as preventive measure.

Mislead a Hacker's Investigations

One opportunity that could be easily overlooked is that hackers who are targeting a certain target rely heavily on their initial investigations of that target. They do reconnaissance and footprinting to get a map of the IP addresses at a site and to try to understand the functions and vulnerabilities of the servers at those addresses. They map DNS names to IP addresses and vice versa to get the information they need.[9]

There are various ways to strike back at the tools hackers use:

(1) Vulnerability scanners make use of service banners. Banners can be falsified or even embedded with malicious strings that contain commands. For example, if the hacker is putting results into a database, a SQL injection exploit could be used.

(2) Vulnerability scanners also search for vulnerabilities by querying servers and looking for certain responses. Messing with responses can break hacking tools or mislead a hacker.

(3) A spider can be trapped in a "tar pit" by creating bogus links which never lead anywhere.

To target these techniques effectively, we must be able to tell that someone sending requests is an attacker. We can

(1) Check to see if a request matches a known bad signature

(2) Send malicious code that will only affect certain hacking tools.

(3) Send affirmative messages regarding locations that don't exist. Normal users will not be looking for such places.

(4) Garden path links on web pages can be hidden from normal users, but spiders will find them.

Tracking File Transfers

Executables which are not meant to be copied can be embedded with a trojan horse which can announce the IP address and/or personally identifiable information of the person who installs or uses the file, and be used to indict that person. For example, an attacker who was trying to leak files off of a private network could be presented with such a file. After installing, and the trojan phoning home, they could thus be accused of both hacking into a network and of stealing the file.

Moreover, if a hashing scheme like MD5 is used to uniquely identify such a file, and that hashing scheme can be exploited, a trojaned file could be made to look identical to a normal file. MD5 has recently been shown to be vulnerable, such that another different file can be discovered with a duplicate hash in a relatively short time. If two binary sequences which hash to the same value are cleverly placed in the file, the difference between the two files could be used as a boolean switch which turns the trojan off or on.[10]

Fake Vulnerabilities

When an hacker attacks, she often uses a tool such as Metasploit, which not only exploits the vulnerability, but provides the user with a remote command shell executing on the remote computer. Since the victim controls that command shell, it is possible to cleverly provide a shell which will load malicious code onto the attacker's machine if she is not careful.[9]

To pull this off, one must first set up a fake vulnerability. The computer will respond to probes according to how a vulnerable computer would. When the malicious payload is sent, the server responds with a command shell that issues an invisible command (the text color is the same as the background color); all the attacker need do is hit ENTER.

At this time, the attacker's computer can simply be rendered unoperational, or whatever information to be discovered can be discovered. Presumably the most careful attackers would not allow themselves to be identified by such an attack, but a botnet could be analyzed similar to how it is done with a honeypot.

Distributed Denial of Service

In this scheme, users agree to take part of a DoS attack against a known malicious site, such as with Make Love Not Spam.[5] It can be done while the users are idle so as not to eat up their extra cycles; for example, with a screensaver.

V. Feasibility

General Observations

Costs and Benefits

We make a few general observations about the costs and benefits of counter-attacks. First, the costs of developing counter-attacks are ongoing. One must continuously develop new techniques, because once a technique has been used in practice, attackers will learn to work around it. Computer security is an arms race.

Second, a counter-attack produces no long-term benefits. It eliminates the immediate threat, but it does not solve the underlying problems of poorly-administered PC’s and inadequate investment in security. Vulnerable hosts that are cleaned up will soon be re-infected. The only way that counter-attacks could change this is by punishing negligent users; but this cannot be justified as self-defense. (Moreover, if we wanted to encourage users to pay more attention to security, attacking their machines is probably not the best way to do it.)

In fact, counter-attacks may cause long-term harm, because they introduce new malware into the wild. Hackers already know how to reverse-engineer software patches to discover vulnerabilities, and it is natural that they would take counter-attack agents and adapt them for their own use. The danger is that we might give attackers a better weapon than they could build on their own; a counter-attack agent should be more robust and reliable than a wild worm.

The decision of whether to deploy a counter-attack resembles the decision of whether to disclose a software vulnerability.[11] A counter-attack has obvious benefits, but one must judge whether it gives an even greater advantage to the enemy. We currently know very little about this trade-off.

Escalating the Conflict

What actions might an attacker take, when confronted with a counter-attack? The attacker might return using more aggressive intrusion techniques. The attacker might develop new malware that is harder to remove. For instance, a worm might threaten to destroy its host computer if tampered with; effectively turning the host into a "human shield," and using the threat of liability to deter a counter-attack. These scenarios are all speculative, but one should keep them in mind when considering the consequences of attacks and counter-attacks.

Offense is Easier than Defense

In spite of the uncertainties about how a counter-attack would play out in the real world, offensive tactics do seem to have an advantage over defensive ones. The basic problem is that "bad guys can attack anywhere, good guys have to defend everywhere." To put it in more concrete terms, an attacker succeeds if he finds a vulnerability, whereas a defender succeeds only if he finds the same vulnerability as the attacker.[12] Thus, attacking the attacker is likely to succeed. What we don't know, is whether attacking the attacker will actually make us any more secure.

Suppose, for the sake of discussion, that counter-attacks do provide a net increase in security. Then counter-attacks have an interesting property: overall security might no longer depend on the weakest link (i.e., the minimum effort put forth by any individual host on the network). If a few hosts launch aggressive counter-attacks, this could deter attackers and improve security for everyone. Overall security might depend on the total or average effort of all hosts on the network. This would have implications for the design of economic incentives to improve information security.[13]

Anti-worms

This counterattack is technically feasible to engineer and deploy.[7,8] It is also economical to develop, though, it's not clear how quickly an anti-worm could be developed to neutralize a worm when it is spreading its fastest and doing the most damage.

One interesting possibility is that if there is more than one independent anti-worm, the anti-worms might attack each other.

This attack depends on being able to categorize the worm on the offending machine. Certainly an antiworm could stop the spread of a known worm in this way. The antiworm could also try a series of attempts against a worm whose behavior is less well-known. Most likely to work would be to try blocking various ports.

Note that if a worm sends packets with spoofed source addresses, it cannot be targeted by an antiworm. However, some ISP's do egress filtering, which drops spoofed packets. Also, a worm that uses a connection-based protocol such as TCP cannot spoof packets.

An ISP would be a good candidate to provide intrusion detection systems that retaliate with anti-worms. It is in an ISP's best interest to restrict malicious activity from within. It could neutralize botnets among its own clients, and potentially even inform the client that his machine was corrupted and subsequently neutralized, and provide steps to fix the vulnerability. However, setting up such a server is difficult and expensive and requires frequent update. For this type of technique to gain widespread feasibility, software would need to be sold to many ISPs.

Spam

Email based counter attack is quite feasible using spam detection devices and configuration of the Email servers. But counter attack gets directed most of the time at innocent servers due to sender address spoofing. Another concern is the network traffic generated from the bounced emails and unsolicited messages to administrators. Also note that if spam/virus filters are too sensitive, the system may trigger a counter attack even for legitimate mails. Having an understanding between Email Server Admins, the notification task overhead and liability resulting from the action can be greatly reduced.

Botnet

The counter attack on a Botnet attacker depends heavily on the fact that we are able to monitor the communication between the compromised host and Home machine. Traps like Honeypots and Honeynets[14] can be used lure the attacker. Being able to monitor the command channel would also help in collecting evidence for legal purposes.

Mislead a Hacker's Investigations

A fundamental feature of this interaction of a hacker querying a server is that a server can control what information the attacker receives; that information need not be truthful or even harmless. Another feature is that hackers largely use a set of known tools to automate this investigation; such known tools can be manipulated. Thus, it can be quite easy to frustrate a hacker.

Care must be taken that normal users aren't affected by these efforts, but this is often possible. These countermeasures can be expensive to implement or cheap, depending on the complexity. It seems likely that a dedicated hacker who was being paid to attack a site would eventually be able to get around such measures, but these would likely buy enough time for the site so that it could take measures to secure itself or, say, remove important information, before the attack succeeded. With some of the countermeasures, the attacker would run the risk of being traced before she completed her task.

These measures would be best suited to be used selectively by companies who present a sensitive web target. A full understanding of the internet face of the company would be needed to aid effective implementation, so I foresee little in the way of a packaged security solution in this area.

Tracking File Transfers

Tracking what happens to a file after it leaves your network can be essential for indicting a hacker for an attack or for stealing protected information, and also those who copy the file from the hacker. With a hash key that is identical to a non-trojan file, this could be very difficult to detect if the file is distributed over a peer-to-peer network.

Thus, this counterattack does not defend or divert an attack, but it can serve to aid the followup investigation when an attacker does break through. It can also quell the spread of a file if people become aware that they may have trojans. It is important to have methods which not only deter but can actually harm the attacker; otherwise, the attacker has infinite ability to play with different ideas. There are many hackers that would not do so if they thought there were more risk of being caught.

This exploit is easiest to devise on executable files, most notably installer files, but it is conceivable that it could work on data as well, if that data can exploit the executable to which it is inputted.

Fake Vulnerabilities

Most hacking attempts occur against widely known vulnerabilities, such as after a patch has been released. Hackers first attempt to scan for known vulnerabilities before trying more outlandish things. Thus, fakes can be set up for any known vulnerabilities, and most attacks to a site will have an opportunity to be taken in. Hopefully one can thwart the attacker before a real vulnerability is discovered!

It is not clear how likely a hacker is to hit "ENTER" when presented with a nefarious command shell.

A successful counterattack could easily bring down the attacker's computer, and in some more rare cases may help to investigate an attacker and bring them to justice. For example, it could detect transmissions between the computer and a base computer, presumably more likely to be one that will identify the attacker. Unfortunately, this counterattack is likely to affect an innocent user's computer that has been coopted by the attacker, so care must be taken.

Distributed Denial of Service

This technique was a proven "success" with Lycos Europe, in the sense that it was popular enough to do effective DoS attacks. Other distributed computing efforts like setiathome have also been successful; people are more than willing to give their free cycles. There are tools such as SpamCop which track websites which use spam to advertise themselves (or 'spamvertisers'), so the chance of directly harming an innocent is extremely low. Other organized cybercrime websites could be attacked as well. The effect of this attack is to increase the cost of spamming or other cyber crimes by reducing the bandwidth available to the sites and increasing their bandwidth usage. These sites are often charged by ISPs for their bandwidth usage.

However, this attack is scary because of the amount of overall internet bandwidth it could eat up. It could perhaps affect the quality of life for random internet users who shared subnets with target sites, though to what extent we did not calculate.

It would also be necessary for a well-trusted institution to initiate such an attack, or there would be too much chance of the DoS being used for ill. Such an institution would open itself up to considerable criticism, as did Lycos Europe.

VI. Legal issues

Cyber counter-attacks occupy a legal gray area. A business that launches a counter-attack may violate the same criminal laws as the attacker, and may bear liability for damages resulting from the counter-attack. On the other hand, a counter-attack may be justified as “self-defense” or “self-help,” if the method used is proportional to the initial attack, and there were no reasonable alternatives. The situation is complicated by the fact that cyber-attacks are novel crimes, without any analogue in the real (non-computer) world. Any doctrine of self-defense in cyberspace must take these new circumstances into account.

Background: Computer Crimes

We briefly summarize some sections of criminal law that apply to attacks and counter-attacks. There have to date been no legal proceedings against counterattackers, so we are forced to speculate.

There is a loose distinction between computer-related crimes (ordinary crimes committed with the aid of computers) and true computer crimes (crimes that are only possible using computers). The kinds of cyber-attacks we are considering tend to be true computer crimes, in that they exploit the unique properties of the Internet, such as anonymity and global connectivity. Traditional common law concepts break down when we apply them to these cases. "Property" and "theft" are not clearly defined in this setting; there is only information, which can be copied or altered. "Trespass" also loses its meaning for some attacks; any host can communicate with any other host on the Internet, and it is the type of communication (an e-mail versus a denial-of-service attack) that matters.[15]

Instead we must rely on specialized computer crime statutes. At the federal level, the statute most relevant to cyber-attacks is the Computer Fraud and Abuse Act (CFAA). Broadly speaking, the CFAA makes it a crime to "intentionally access a computer without authorization" in order to pursue certain kinds of criminal conduct.[16] Though originally intended to combat fraud, the CFAA has since been amended to deal with viruses, worms and other "malicious code." It punishes anyone who "knowingly causes the transmission of a program, information, code, or command," and as a result, "intentionally causes damage without authorization" (a)(5).[13] Note that two elements are required: unauthorized access, and intent to cause damage.[12] So a "neutralizing" counter-attack would not automatically violate this statute. Consider a procedure that disables a piece of malware running on an innocent third party's computer, but does not touch anything else. This might constitute an unauthorized access, but it is not a violation because there is no intent to cause damage.

Neutralizing counter-attacks may be allowed under the CFAA, because the statute rates the severity of a crime according to the amount of (economic) harm it causes. However, there is an important caveat: the monetary cost of a computer intrusion depends heavily on the actions of the victim.[17] If the victim feels that the incident has hurt its reputation, it may carry out a lengthy forensic investigation and extensive security upgrades, which effectively increases the cost. On the other hand, if the victim urgently needs to get the compromised system back online, it will spend much less on investigation and remediation. Furthermore, there may be intangible harms such as loss of privacy or access to confidential information; assigning these a monetary value is subjective at best. So there is always a danger that a benign counter-attack could trigger an expensive over-reaction by the victim, which would make it into a crime.

In addition, many states have passed computer crime laws, which are usually more aggressive and up-to-date than the federal CFAA. California law, for instance, punishes anyone who "knowingly accesses and without permission adds, alters, damages, deletes, or destroys any data, computer software, or computer programs" (c)(4).[18] This statute does not consider whether the intent is to cause damage, or whether the data being altered is itself legitimate (as opposed to malware); all kinds of attacks and counter-attacks are illegal. However, there is an exemption for "acts which are committed by a person within the scope of his or her lawful employment" (h)(1).[15]

Self Defense

Self defense ordinarily applies when a person is threatened with serious bodily harm. It requires two elements: "a counterstrike ... which is proportional" to the threat, and "a good faith, objectively reasonable belief that the counterstrike was necessary, in the sense that there were no adequate alternatives".[19] We remark that while the basic principle is easy enough to understand, applying it in practice can be a subjective matter. For example, self defense usually means that one has the right to "stand one's ground," i.e., backing away does not count as an adequate alternative. But should we apply this standard when the counterstrike involves deadly force? Or should we permit the use of deadly force only as a last resort, when one cannot retreat? The debate over a recent gun law in Florida shows that, even in the real world, self defense is sometimes controversial.[20] We encounter similar issues when we try to apply self defense in cyberspace.

Claiming self-defense in the case of a cyber-attack raises a number of issues. First, it is difficult to estimate the danger posed by an attack that has not yet occurred. As an example, one might want to interfere with a hacker who is scanning the network for vulnerabilities. Self defense would not apply in this case, unless one can show that an attack is imminent, and that one's actions are proportional to the magnitude of the threat. Suspicious activity alone does not justify a preemptive counterstrike (although one could still use nondestructive techniques to confuse or distract the hacker).

Another problem is that the attack and counter-attack may involve an innocent third party. Consider a computer intrusion that is routed through several compromised hosts, or a DDOS attack using a botnet. Since the attack is carried out using machines that belong to an innocent third party, the counter-attack may affect the third party more severely than the attacker. This is troubling; one's intuitive notion of self defense suggests that the counter-attack should be aimed primarily at the attacker, and should not result in excessive amounts of collateral damage. This goes beyond the usual requirement of proportionality.

Lastly, in the case of a cyber-attack, it may be difficult to show that there were no reasonable alternatives to a counter-attack. This is a serious problem -- the choice of which action to take is complex, and a network administrator may have a very different perspective from a judge or jury.[16] Also, whether a particular alternative is "reasonable" may be a matter of opinion. In many cases, a counter-attack is not the only option, but merely the cheapest. One can always buy more bandwidth and servers to withstand DDOS attacks, and deploy firewalls and packet filters to stop worms. The main obstacle here is cost. On the other hand, money is precisely the objective of most cyber-crimes. What level of expense is constitutes a "reasonable" alternative to a counter-attack? The only way this question can be answered is if the computer industry comes to an agreement on what the trade-offs should be.

Other Considerations

Is Computer Intrusion a Police Power? The laws on computer intrusions are much less developed than the laws that govern the use of physical force in the real world. To a large extent, the computer industry is left to manage security as it sees fit. This may change if attacks and counter-attacks become a major threat. Just as in the real world, where only police are allowed to obtain a warrant and break into a home, a parallel might develop on the internet, where certain parties will be licensed to break into a computer and neutralize threats. This right could be granted by a government, or perhaps by agreement between an ISP and its clients.

Evidence Obtained by Counterattack: If a counterattack succeeds in discovering the identity of the hacker, perhaps by attacking their machine and discovering personal information, is that information admissible in court? This will likely be a defensive tactic used by any hacker. Currently such evidence is not admissible. However, some countries have proposed a different standard, where the severity of the crime being prosecuted is balanced against the severity of the hack used to obtain the evidence.[21]

Jurisdiction: If the source and target of a cyber-attack are located in different states or countries, the action falls under both jurisdictions.[12] Thus one must consider the consequences of a counter-attack under other nations' laws. This is especially true for multinational corporations, because their subsidiaries in foreign countries could be affected. International law enforcement is often weak, but foreign governments might also respond through diplomatic or even military means. This has to be considered on a case-by-case basis.[22]

Civil liability: Even if a counter-attack is completely legal, it may result in civil liability. In the absence of clear industry standards regarding security, courts have to devise their own standards. Courts have been assertive in assigning liability in these cases.[23]

Specific Scenarios

Anti-worms: These are probably legal, as long as the anti-worm is benign (does not crash machines or cause network congestion). This can be justified using nuisance law, which allows for self-help as long as it does not cause any “breach of the peace” or "unnecessary injury."[16] Also, there are no obvious better alternatives for dealing with widespread worm infections.

Spam: This may be justified using nuisance law, as long as the counter-attack is benign. Self-defense is problematic, because the counter-attack targets third party mail servers, not the attacker who originally sent the spam e-mails.

Botnet: Similar considerations anti-worms.

Misleading hackers: The nondestructive techniques presented are probably legal. Techniques that cause damage can possibly be justified as 'self defense,' if one can prove that the hacker intends to launch an attack, but this is shaky at best in the current legal framework. Arguments about proportionality and necessity lead into unexplored legal territory.

Fake Vulnerabilities: In this scenario, it is easily proven that any retaliation is in self-defense, but there is a large issue of collateral damage, as one may also bear liability for damaging third party computers that the hacker was using. There is little threat of litigation from the hacker himself, who would be faced with a counter-suit if he filed suit. But if the nefarious shell caused collateral damage, a third party figured out what went on it could raise charges for damage to its machine. In practice, the third party, who had a vulnerability in the first place, is unlikely to know enough to raise suit. Regardless of the legal threat, the most ethical actions are thus those that could bring down only the attacker's applications or help discover the identity of the hacker, with no collateral damage.

Tracking file transfers: This is probably legal, for files which are only intended to circulate within an organization. Tracking files which are legally distributed to the public (e.g., music CD's) may violate privacy laws.

Reverse DDOS: This is almost certainly illegal. A DDOS attack can hurt innocent web sites that are virtual-hosted on the same IP address as the target site, and it can create congestion that affects all hosts on the network. It is not clear what possible threat could justify this kind of reponse, in terms of proportionality or necessity. Lycos's "Make Love Not Spam" drew a lot of fire as an especially flagrant example, which it defended, as mentioned above, by saying it wasn't really a DoS attack, which are illegal in the US and most European states.[5] It is worthy of note that, to our knowledge, legal action was not taken against this effort.

This method could conceivably be legal for cybercrime enforcers who are licensed to do it, but it would likely take a drastic internet threat to warrant such an action.

VII. Recommendations for the future

Improving cooperation on security issues

Cybersecurity requires cooperation among all the players -- users, ISP's, software developers and security providers. There is no "least cost avoider"; security depends on actions at all levels, or "defense in depth." This applies equally to the use of offensive or defensive tactics. Counter-attacks work best when there is at least some coordination among the players; they are most dangerous when an individual end-host takes unilateral action against another host. When an attack occurs, all players should consult with each other about what possible actions they can take. Ideally, one should launch a counter-attack because it is the best option, not because it is the only one available.

The computer industry does seem to be trending towards greater cooperation. ISP's employ a variety of security measures, such as egress filtering of spoofed packets, and the use of specialized devices to "scrub" traffic streams of malicious packets.[24] Also, some ISP's are providing security services for customers, such as "black-holing" DDOS attack traffic, and coordinating with the customer to respond to attacks.[19] The driving force behind this appears to be the threat of liability (for example, the Sarbanes-Oxley Act).[19]

Attribution for crime deterrence and effective counter attack:

Correct and fast attribution of the crime is the key to stopping cyber attacks and making counter-attacks successful and just.

As with many crimes, it’s essential to assign liability to those parties that can take actions to discourage attacks. For example an ISP can he held liable for the damage its network caused in the propagation of a worm or virus. This in turn can trigger following changes:

1.The ISP starts a process of authenticating its subscribers and putting the infrastructure to locate crime originators. This may encourage ISP to run egress filters to avoid sending illegal packets on the network.

2. A cyber-citizen-credit-score rating system may emerge that will allow good cyber-citizens to subscribe to the net cheaper than bad citizens. Good Cyber-citizens are people who keep their machines up to date and pay the ISP for keeping it up-to-date.

3. Put together a centralized infrastructure for counter-attack as described below.

4. Since attribution of the attack is possible, counter attack can be targeted at the right machines.

5. New network technologies will be developed with security as one of the main design considerations.

Centralized infrastructure for counter attack:

Since the Internet is a shared resource, i.e., a public good, one might consider creating a non profit or government organization that acts as police on the network. When a system detects that it’s under attack, it (with or without consent of the user) informs the Cyber police. The Cyber police machine equipped with all the necessary tools does the authentication of the complaint, then either neutralizes the attacking party or if required deploys a counter attack. While the attack is carried out it collects all required logs for future investigations. This just means of centralizing the counter attack strategy to reduce cost and provide neutral intervention. Cyber Police has to be a neutral party, equipped with the best equipment and strategy. Cyber police and the procedure it will follows have to be open. But this also means that it’s vulnerable to attack.

The obvious open question is how much authority the Cyber Police should have. One disadvantage of this arrangement is that it may interfere with the development of new technologies on the Internet. Also, a centralized authority may not work well when different users have different security needs; for instance, some users may place a greater value on bringing a hacked machine back on-line, than on prosecuting the attacker.

VIII. Conclusion

Under a rather loose definition of cybercrime, according to Valerie McNiven, at $105B, total annual revenue for cybercriminals in the world is higher than for drug trafficking.[25]

It remains to be seen what shape counterattacks will take, but under the mounting pressure to do something about it! their continued emergence is inevitable, as statistics pour out about how much cybercrime costs the world. Hackers have been able to have free run of the internet, only being brought in if they make a silly mistake. Adequate countermeasures will not only help deter attacks when they happen, but more importantly, deter hackers from launching them in the first place.

It will take years for legal systems to set preference and formalize their thoughts on these matters, and equally long or longer for inter-jurisdiction litigation to proceed effectively and at more than a crawl.

Until then, our hope is that brave and angry souls will be willing to light the right fires to counter the hackers and continue to raise questions about what is cost-feasible, ethical, and legal.

[1] Niall McKay, "Pentagon Deflects Web Assault," *Wired News* (September 10, 1998)
[2] Landergren, "Hacker Vigilantes Strike Back," cnn.com (June 20, 2001)
[3] Majik, "Code Green. Are you Serious?!", http://www.xatrix.org/article.php?s=684 (September 6, 2001)
[4] "Symantec Security Response - W32.Welchia.Worm," http://securityresponse.symantec.com/avcenter/venc/data/w32.welchia.worm.html
[5] David Dittrich, "How bad an idea was 'Make Love Not Spam?' Let me count the ways." http://staff.washington.edu/dittrich/arc/workshop/lycos-response-v3.txt
[6] Symbiot, "On the Rules of Engagement," http://www.symbiot.com/riskmetricsdownloads.html
[7] Frank Castaneda and et al, "WORM vs. WORM: Preliminary Study of an Active CounterAttack Mechanism," Workshop on Rapid Malcode (WORM) 2004.
[8] Timothy M. Mullen, "Defending Your Right to Defend," in Advanced Network Self Defense, Syngress, 2005. Timothy M. Mullen, "Enforcer," Black Hat Windows Security 2003. Laurent Oudot, "Fighting Internet Worms With Honeypots," http://www.securityfocus.com/print/infocus/1740 (2003).

[9] H. Meer, R. Temmingh, and C. van der Walt, "When the Tables Turn:Passive Strike-Back," in Advanced Network Self Defense, Syngress, 2005.

[10] Dan Kaminsky, "MD5 to Be Considered Harmful Someday," in Advanced Network Self Defense, Syngress, 2005.

[11] Eric Rescorla, “Security holes... Who cares?” Proceedings of the 12th USENIX Security Conference, August 2003.

[12] Ross Anderson, “Why Information Security is Hard – An Economic Perspective,” http://www.cl.cam.ac.uk/ftp/users/rja14/econ.pdf

[13] Hal Varian, “System Reliability and Free Riding,” http://www.sims.berkeley.edu/~hal/Papers/2004/reliability

[14] The Honeynet Project & Research Alliance. http://www.honeynet.org/papers/bots/

[15] Rasch, Mark D. “Criminal Law and the Internet.” Book chapter in The Internet and Business: A Lawyer’s Guide to the Emerging Legal Issues, Joseph F. Ruh Jr. (ed.), 1996.

[16] Title 18, U.S. Code, Sec. 1030.

[17] Granick, Jennifer S. “Faking It: Calculating Loss in Computer Crime Sentencing.” Workshop on Economics of Information Security (WEIS) 2005.

[18] California Penal Code, Sec. 502.

[19] Karnow, Curtis E.A. “Strike and Counterstrike: The Law on Automated Intrusions and Striking Back.” BlackHat Windows Security 2003, Seattle, WA, Feb 27, 2003.

17 Roig-Franzia, Manuel. “Florida Gun Law to Expand Leeway for Self-Defense.” The Washington Post, Apr 26, 2005.

[21] Stephen Bell, "Hacker evidence admissible in court?" Computerworld New Zealand, Nov. 26, 2004.

[22] "Agora Active Defense Workshop Report," Sept. 12, 2003, http://staff.washington.edu/dittrich/ad/

[23] William Cook and Wildman Harrold, "Put Some Bite in Your Information Technology Protection, or It Can Bite You Back," FindLaw Modern Practice, Jan. 2005.

[24] Stefan Savage, e-mail communication.

[25] Souhail Karam (Reuters), "Cybercrime pays off more than drug trafficking, security expert says." http://www.computerworld.com/securitytopics/security/story/0,10801,106574,00.html, Nov 28, 2005.