A Survey of Challenges in Attribution

W. Earl Boebert
Sandia National Laboratories (retired)

1 STATEMENT OF WORK

This paper was prepared under the following Statement of Work:

The consultant will research, draft, present, and finalize a commissioned paper on the problem of attributing cyber intrusions, both technical (can you determine which machine or machines are generating or controlling the intrusion?) and nontechnical (can you determine the party that should be held responsible for the intrusion?). Given that the "attribution problem" is seen as a major barrier to implementing any national policy to deter cyberattacks, this paper would lay out the technical and nontechnical barriers to attributing cyberintrusions, explore plausible aspirations for how these barriers may be overcome or be addressed in the future, and describe how and to what extent even perfect technical attribution would help to support national policy for deterring serious cyberattacks against the United States (that is, cyberattacks with a disabling or a crippling effect on critical societal functions on a national scale (e.g., military mission readiness, air traffic control, financial services, provision of electric power)).

2 INTERNET ATTACKS

2.1 Structure of the Internet

The Internet is a packet-switched network. In simplified form, such a network consists of interconnected nodes, each of which is given a number called an IP address. Units of data to be transmitted are cut up into pieces, or packets. Each packet is given a header which contains (in this simplified discussion) the source and destination IP address of the transmission and other control information. Packets are sent to routers which are provided with tables that give a "next reasonable router" for any given IP address. This relieves the routers from having to know all possible routes to all possible destinations. To use a geographic analogy, consider a router in Albuquerque that wishes to send a data element to New York. The element will be cut into packets whose source IP address will designate Albuquerque and destination IP address will designate New York.

To initiate transmission, the router in Albuquerque only needs to know the IP address of some intermediate router in the general direction, for example, Kansas City. That router, in turn, needs only to know that St. Louis is "on the way" and so forth, "hop" by "hop," until the destination is reached

and the packet is joined with others to reconstitute the data element. The dynamic assignment of routes at each hop is what gives packet-switched networks their great resistance to failure.

Early in development of packet-switched networks it became obvious that numeric IP addresses were difficult for humans to cope with, and a second facility, called the Domain Name System or DNS, was developed to permit use of the symbolic domain names with which we are all familiar. The DNS is a distributed lookup system which, when queried, converts a symbolic name for a node, called a "hostname," to a numeric IP address.

2.2 The Nature of Attacks

The Internet shares the central vulnerability of all cyber systems: every action initiated by a human is performed indirectly by system functionality which cannot be directly observed. This vulnerability can be exploited by the insertion of malicious functionality in a node. Since the attacks take place on a packet-switched network, the malicious functionality that carries them out will, in general, be distributed over multiple nodes which must communicate over that network. Thus there are two basic elements to Internet attacks: malicious functionality and malicious packets. Malicious packets are those which are either used to communicate with a malicious functionality or used to shut down a target node by flooding it with traffic that it cannot handle.

If the malicious packets are part of a flooding attack, then the source IP address can be forged because no return communication is desired and attribution is not available from the packets themselves. If the hostile packets are control packets then technical attribution is at least theoretically possible, because they will contain a source IP address to permit return communication, as in an interactive session.

The well-publicized attacks on Estonia[1] and the nation of Georgia[2] were flooding attacks directed at the public interfaces of government services provided over the Internet. Such attacks are relatively easy to mount and very difficult to attribute, especially if they involve so-called "botnets" as described below. Major attacks against significant societal functions such as the military or power distribution require that the attacker gain access to internal network resources and not just direct streams of packets at public interfaces. Consider, for example, an electric power utility. If the utility were to be attacked the way Estonia and Georgia were, customers would not be able to pay their bills on line but power flow would not be affected. To disrupt power the attacker must penetrate the utility's security perimeter and be able to manipulate and destroy data and functionality on internal control networks. A similar situation exists with regard to the military, financial systems, and other major societal services in that their critical functions are not dependent upon the operation of public nodes such as World Wide Web sites.

It is unlikely that an adversary will initiate an attack on critical internal networks without first conducting both external and internal reconnaissance of the target network. Internal reconnaissance will generally be aimed at collecting "metadata," the internal information such as router tables, internal IP addresses and so forth which the network uses to manage itself. This data will be analyzed to determine network topography as well as what actions during the actual attack may produce the greatest degree of disruption to the service being provided by the target. The analogy here is with a sabotage team scouting a building to learn where to place their explosive charges for maximum effect.

The reconnaissance phase is the period where the attacker is most vulnerable to technical attribution. Since the attacker is, in effect, on a voyage of discovery through largely uncharted waters, it not only must exfiltrate large quantities of data but also typically must make repeated intrusions as its analysis raises questions that must be answered. A successful disruptive attack on a major network is a considerably more complex undertaking than flying an airplane into a building or setting off a truck

[1]Myers, Steven Lee. "Cyberattack on Estonia stirs fear of 'virtual war.'" *New York Times*, May 18, 2007. Available at http://www.nytimes.com/2007/05/18/world/europe/18iht-estonia.4.5774234.html. Accessed May 31, 2010.

[2]Anon. "War, redefined." *Los Angeles Times*, August 17, 2008. Available at http://articles.latimes.com/2008/aug/17/opinion/ed-cyberwar17. Accessed May 31, 2010.

full of explosives in a crowded street, and considerable planning is required to raise the probability of success to a point where a rational decision to proceed can be made.

2.3 The Attribution Problem

The general problem of attribution can be broken down into two subsidiary problems: technical attribution and human attribution. Technical attribution consists of analyzing malicious functionality and malicious packets, and using the results of the analysis to locate the node which initiated, or is controlling, the attack.

Human attribution consists of taking the results of technical attribution and combining it with other information to identify the person or organization responsible for the attack. There are significant barriers to both forms of attribution.

Both forms of attribution can be either positive or negative. Positive attribution yields positive information, such as where a node is or the identity or other characteristics possessed by an associated human. Negative attribution yields specific information, such as where the node is not located, or that the citizenship of an individual is not that of the U.S. or its allies. Negative attribution is useful in determining courses of action, such as whether suppressing a malicious node is a domestic law enforcement or a foreign covert action problem.

Both technical attribution and human attribution have degrees of uncertainty associated with them. One may know the country of registration of a machine, but not know the current physical location because the machine may be portable and accessing the Internet over long-distance telecommunication. Likewise, one may know the identity of attacker (e.g., through an informant or other "off net" information such as wiretap) but not know that person's affiliation or sponsorship, knowledge of which is important in determining what kind of retaliatory action is appropriate.

3 BARRIERS TO FORENSIC-BASED TECHNICAL ATTRIBUTION

3.1 Overview

The barriers described below represent impediments to conventional, forensic analysis, mostly of hostile packets retrieved from logs or captured by real-time monitoring on the Internet as it currently exists. This analysis, uncertain as it is, is only possible when administrators are alerted to a reconnaissance activity; in a large-scale disruptive attack, the first information to be destroyed by any competent attacker will be logs and other data which may facilitate attribution. There is, however, the possibility of post-attack analysis of logs saved in a protected fashion, such as off-site backups, with the aim of discovering the traces left by pre-attack reconnaissance.

Analysis of malicious functionality extracted from compromised machines is typically of greater value in determining the capabilities of the attacker and likely nature of the possible upcoming disruptive attack than it is in determining attribution. The amount of information on the Internet about malicious functionality is so large that a relatively low level of technical competence is required to exploit it. In the case of malicious software, it is not, as is the case with physical objects, to assert that a particular level of sophistication is an indication that a national laboratory produced the device. Brilliance in software development can be found anywhere, and the only physical resources required are a laptop and an Internet connection.

3.2 Botnets

A relatively new phenomenon is that of the botnet, which is a very large collection of nodes containing malicious functionality and operated under centralized control. Estimates of botnet sizes are highly

uncertain and range from tens of thousands to millions of machines per net.[3] The principal use of botnets is the sending of spam email and flooding attacks called "DDoS" for "distributed denial of service."

Botnets are constructed by mechanisms which exploit vulnerabilities in operating systems or third party software to gain administrative privilege and install malicious functionality. These mechanisms can be self-propagating ("viruses" or "worms"), arrive by email ("phishing") or they exploit vulnerabilities in the DNS system. In these mechanisms a query for the IP address of some legitimate site actually returns the IP address of a malicious site that appears to be the legitimate one but really exists to insert malicious functionality in the visiting node.

There are two classes of victims of botnets. The first is the target of the coordinated attack; the second is the group of individuals whose machines have been captured by the botnet operator. In earlier virus and worm attacks, these two classes were identical, that is, the machine being infected by the virus or worm was also the target. This is not the case with botnets; the owner of the infected machine may not know nor care that some percentage of that machine's cycles and Internet bandwidth is being used to attack a third party. As a consequence, there is substantially less motivation for owners of infected machines to incur the expense and risk of removing the malicious software. In response, there have been calls to require the quarantining of machines that have shown signs of having been subverted.[4]

Botnets seriously complicate the technical attribution problem. Prior to their existence, attacks were generally mounted from single nodes. This meant that strings of malicious packets, such as those probing for misconfigured interfaces (so-called "port scans") could be readily correlated to form a picture of the actual attack, which in turn could be used for technical attribution or to guide a response such as discarding packets from that address.

In a botnet attack, the overall attack is distributed over tens to thousands of machines, all with different source IP addresses, thus seriously complicating the correlation problem. The resulting risk is most severe in the area of password guessing. Prior to botnets, password portals could be arranged so that some small number of incorrect password submissions from the same source IP address would cause the connection to be broken. Now one can encounter thousands of attempts at guessing a password, each from a different source IP address, and all coordinated by a single control node. If the attacker is patient and spaces the attempts out at intervals of minutes or even seconds, it may be days between the occurrence of the same source IP address in two different attempts. The attacker may thereby cycle through an arbitrarily large number of possible passwords with little chance of detection.

Botnets are also used for denial of service attacks aimed at shutting down a target node or set of nodes. The simplest method, as described above, is to flood the node with malicious packets. This technique is analogous to jamming in the radio frequency domain. It is most commonly used against World Wide Web sites by less technically capable attackers who object to the site's content or some behavior by the site's owner. More sophisticated botnet attacks involve port scanning or password guessing in order to gain administrator privilege, which in turn is used to corrupt or destroy data, or force shut down, of a critical node.

One technique for achieving technical attribution of a botnet control node is that of so-called "honeypots."[5] These are deliberately vulnerable machines that are placed on the Internet in the hope that they will have malicious botnet functionality installed in them. The malicious functionality, and the honeypot's communication with its control node, are then analyzed to determine the IP address of that control node. As noted below, botnet managers have options to make such analysis difficult.

---

[3]Daniel Woolls, "Spain: Mastermind of 'botnet' scam a mystery" *San Jose Mercury News*, March 3, 2010. Available at http://www.mercurynews.com/ci_14504717. Accessed 31 May 2010.

[4]Mason Rice, Jonathan Butts, Robert Miller, Sujeet Shenoi, An analysis of the legality of government-mandated computer inoculations, International Journal of Critical Infrastructure Protection, Volume 3, Issue 1, May 2010, Pages 5-15, ISSN 1874-5482, DOI: 10.1016/j.ijcip.2010.02.002. (http://www.sciencedirect.com/science/article/B8JGJ-4YC811B-2/2/790562795030f1a318fecc4c5bba0463). Accessed July 9, 2010.

[5]http://www.honeynet.org/. Accessed 31 May 2010.

3.3 Registration Privacy

In order for a symbolic hostname to be recognized by the DNS, it must first be registered by an organization authorized to do so, called a registrar. The registration must include a name, address, telephone number and email for an authorized point of contact for the host. In the early days of the Internet (and its predecessor, the Arpanet) this information was used for cooperation between administrators when network problems arose. The information was, historically, publicly available through a facility called "whois." A common forensic step, when confronted with a suspect packet, was to examine the source IP address, do a reverse DNS lookup to obtain the hostname, and then do a "whois" to obtain basic information such as country of origin. The usefulness of this step has been diluted in recent years with the rise of privacy protections on "whois" data. In some cases the protections are imposed by the registrars; in other cases, a proxy service is provided in which the hostname is registered by the service and the actual owner is not even known to the registrar. The steps needed to obtain the true "whois" information vary from service to service, and can range from simple telephone or letter requests to subpoena.

3.4 Proxies

Proxies are intermediate nodes that perform technical services during a transmission, such as caching packets to improve performance. Proxies, in general, complicate attribution because they change the source IP address of a packet from that of the actual sender to their own address in the course of performing their service. Two classes of proxies with particular impact on attribution are Network Address Translators and Anonymizing Proxy Servers.

3.4.1 Network Address Translation

The growing popularity of the Internet in the 1990s led to a shortage of IP addresses. One response to this was the development of a range of technologies called Network Address Translation or NAT. When IP addresses were first standardized, some numbers were reserved for private networks. An NAT "hides" a private network behind a public gateway. A variety of techniques are used to insure that two-way communication through the gateway is implemented without confusion. By this means large institutions with thousands or hundreds of thousands of machines connected to the Internet may display only a few hundred IP addresses to the public Internet. As a consequence, forensic examination of hostile packets may reveal a source IP address that indicates only the major institution from which the packet came. More detailed attribution requires cooperation of the institution, which may either be impossible (owing to absence of detailed logs) or not forthcoming. The latter is often the case when the packet in question has crossed national boundaries.

3.4.2 Anonymizing Proxy Servers

Other proxy servers are dedicated solely to anonymizing, and many of these are offered as a free public service. They are used by persons desiring privacy on the Internet, and also to bypass location-based content controls imposed by national regimes who seek to deny their citizens access to certain Internet sites. A site which tests and lists public proxy servers[6] documents, at the time of this writing, more that 200 servers in 14 countries.

It is a straightforward process to implement and operate a private anonymizing proxy server, either on one's own machine or one that has been compromised. Many public anonymizing services attempt to limit abuse by restricting their use to access of World Wide Web sites, and therefore would be useful only for external reconnaissance by potential attackers. There would be, of course, no technical limit on the

[6]http://www.publicproxyservers.com/. Accessed May 2010.

services that could be provided by a private anonymizing proxy server. One can easily visualize a botnet in which packets are passed through hundreds or more such servers in a given route, each of which is on a compromised machine with no overt connection to the individual controlling the botnet.

3.5 Dynamic Assignment of IP Addresses

Dynamic assignment is used principally by Internet service providers to minimize the number of IP addresses they need to support their customer base. The mechanism that performs this is called the Dynamic Host Control Protocol or DHCP. In this technique, when a customer machine contacts the provider through some facility such as DSL or television cable, an unused IP address is selected from a pool of addresses reserved for this purpose. This temporary IP address then becomes the source address of packets coming from the customer machine, but only for the duration of the connection, or less. This means that packets originating from a given customer machine may have different source IP addresses, and packets going to it have different destination IP addresses, as time passes.

The DNS service described above generally, but not exclusively, maintains a fixed relationship between symbolic hostname and numeric IP address. If the IP address of a particular machine is dynamically assigned by a protocol such as DHCP then the usual DNS can not be used to "advertise" its symbolic hostname to the Internet. There are, however, sites which provide what are called "dynamic DNS" services. These permit the owner of a hostname to rapidly change the IP address associated with that name, and do so in coordination with DHCP.

Dynamic DNS services can be used by botnet controllers to complicate forensics. If the malicious software in a single "bot" contained the IP address used to communicate with its controlling node, then that address could be used either to physically locate the node or to attack it over the net. Replacing an IP address in the "bot" software with a symbolic hostname provides no protection if that hostname refers to a fixed IP address, for then a forensic analyst could simply perform a DNS query and obtain that value. Instead, the dynamic DNS service is used to present the analyst with an ever-changing array of IP addresses with the intent of insuring that when and if a control node is identified it no longer is being used.

3.6 Onion Routing

Onion routing is an extensively documented technique for achieving anonymous communication within a packet-switched network. Implementations vary, but the basic principle is that of a virtual network of "onion routers" that runs on top of a conventional packet-switched network. Each "onion router" has a cryptographic subsystem based on a secret key known only to it and the initiator of the packet.

The initiator chooses a predetermined route through the "onion routers" to the intended destination. This contrasts with the dynamic routing described for the conventional network. Starting with the last "hop" the initiator puts source and destination addresses in the packet and then encrypts both it and the packet contents with the secret of the "onion router" that will initiate the last "hop." The initiator then makes that the contents of a packet, gives it the address of the "onion router" immediately before it, and encrypts it with that onion router's secret key. The initiator works backward in this fashion, router by router, adding a layer of encryption for each hop—hence the "onion" notion.

When the packet reaches a particular onion router, it decrypts it and obtains two things: the address of the next onion router to send it to, and contents that it cannot decipher. In this way compromise or observation of a given onion router yields only information about where the packet is to go next and does not show the source, destination, or contents.

Onion routing is publicly available on the Internet through a free service called "Tor,"[7] maintained by a U.S. registered nonprofit corporation. Volunteers worldwide host Tor routers, and the stated purpose of the service is to preserve privacy on the Internet.

[7]http://www.torproject.org/. Accessed May 2010.

Onion routing is a particularly attractive method for providing anonymous communication amongst elements of a botnet and sufficient documentation on the technique exists to permit implementation by technologists of a range of abilities.

3.7 Covert Communication

Covert communication is the transmission of information using system resources that were not intended for that purpose, i.e., "in a manner surprising to the authorities."[8] An example in the physical domain is the so-called "prison telegraph," where inmates communicate by rapping on the water pipes. In a sense, covert communication is the most powerful anonymizing tool in that it seeks to disguise the fact that communication is taking place at all.

Packet-switched networks in general, and the Internet in particular, provide ample opportunity for covert communication because of the large amount of control information carried in each packet. This information can be used to encode messages to malicious functionality, as can real-time phenomena such as packet sizes, timings, and sequences.

A well-publicized form of covert communication is steganography, in which messages are hidden in other messages or (more typically) pictures. This technique is less attractive as a way to communicate ith malicious functionality because of the overhead associated with extracting the information, but has been used for covert human to human communication.[9]

3.8 Future Indicators

The now almost universal use of digital media, with its ease of copying and transmission, has led to a rise in the unauthorized distribution of copyrighted works, most notably music and motion pictures. The copyright owners have responded to this in two broad areas. The first is the use of so-called "digital rights management," which seeks to impose technical controls on distribution. The second is to take legal action against those involved in unauthorized distribution, both providers and recipients.

In the past the copyright holders had initiated action against large numbers of recipients, such as students and other less powerful individuals in our society. After considerable adverse publicity, the music industry in particular dropped this approach[10] and focused on organizations that facilitate distribution.[11]

Recently, the owners of the copyright to a popular motion picture have resumed the practice of taking legal action against recipients.[12] In this action, as in previous suits by copyright holders against recipients, the plaintiffs have asserted repeatedly that IP addresses can be reliably associated with individuals. Setting aside the technical merits (or lack thereof) of such a claim, if it persists in the legal system then we can anticipate increased activity in the development of tools and techniques to prevent human attribution based on IP address.

In the area of covert communication, there has long been support by the U.S. Government for software that evades attempts by authoritarian regimes to censor the Internet.[13] Research in this area

[8]Morris, Robert, as reported in *Newsletter of the IEEE Technical Committee on Security and Privacy*, Winter 1991. Available at www.list.gmu.edu/misc_pubs/csfw/c91rep.pdf. Accessed May 31, 2010.

[9]Montgomery, David. "Arrests of alleged spies draws attention to long obscure field of steganography," *Washington Post*, June 30, 2010. Available at http://www.washingtonpost.com/wp-dyn/content/article/2010/06/30/AR2010063003108.html. Accessed July 9, 2010.

[10]McBride, Sarah and Smith, Ethan. "Music Industry to Abandon Mass Suits." *Wall St. Journal*, Dec 19, 2008. Available at http://online.wsj.com/article/SB122966038836021137.html. Accessed May 31, 2010.

[11]Plambeck, Joseph. "Court Rules that File Sharing Service Infringed Copyrights." *New York Times*, May 12, 2010. Available at http://www.nytimes.com/2010/05/13/technology/13lime.html?src=me. Accessed May 31, 2010.

[12]Smith, Ethan. "Thousands are Targeted over 'Hurt Locker' Downloads." *Wall St. Journal*, May 29, 2010. Available at http://online.wsj.com/article/SB10001424052748703957604575272843955251262.html. Accessed May 31, 2010.

[13]http://www.dit-inc.us/freegate. Accessed July 10, 2010.

continues,[14] and while the focus of the present tools is on passive observation of blocked Web pages, the basic technology can be applied to the covert communication requirements of attack tools. Whatever the motivation, it should be anticipated that technology that was once the exclusive domain of clandestine services will become more available to ordinary users, whether for good or ill.

4 ELIMINATING BARRIERS TO FORENSICS-BASED TECHNICAL ATTRIBUTION

None of the barriers described above will be particularly easy to eliminate. Proxies, NAT and dynamic assignment of IP addresses are now so deeply ingrained in the structure of the Internet that eliminating them would require significant cost and effort. DHCP and dynamic DNS are fundamental to a democratic Internet which has room for small and specialized servers. Covert communication, because it both exploits needed facilities and is extremely hard to detect, is virtually immune to elimination.

The three remaining cases are the overly privacy-preserving (and therefore attribution-preventing) services of Tor and Anonymizing Proxy Servers and Registration Privacy. Theoretically, all three could be eliminated by legislative action. However, owing to the transnational nature of the Internet, such action would have to be coordinated among multiple jurisdictions, or the service banned in one country would simply move to a "safe haven" where it was legal. Even if Tor as a service was eliminated, the underlying technology is so well known that the creation of a clandestine successor can be predicted with certainty; a similar argument can be made for the continuation of anonymizing proxy servers. If Registration Privacy were somehow to be outlawed, then malicious registrants would simply revert to the practices they used before it was available, and provide false identity, location and contact information on their registration documents.

A new low-level protocol for the Internet, called IPv6, is in the process of being deployed. The principal problem the protocol solves is the upcoming shortage of numeric IP addresses, which is rapidly becoming critical. IPv6 contains many new security features which, if they are properly implemented and administered, promise to reduce the difficulty of technical attribution. The concern amongst the technical community is that first, these features have not yet been subjected to serious attack, second, many of them are optional and may not be invoked by administrators who are under pressure to deliver network performance, and third, the protocol was designed over fifteen years ago and the art and science of attack has progressed significantly.[15]

5 ALTERNATIVES TO FORENSIC-BASED TECHNICAL ATTRIBUTION

5.1 Counterattack

The limitations of forensic-based attribution and the constant increase in the number and sophistication of attacks has led to more serious consideration of the option of counterattack or "hack back" in order to either obtain attribution or suppress attacks. This has reached the point where security researchers are publishing vulnerability information pertaining to the software used to convert an otherwise innocent machine into an element of a botnet.[16] Frustration with the current adverse trends in cyber security may well lead to an increase in "hack back" activity, with or without authorization, with concomitant risk to the overall stability of the Internet.

[14]Burnett, Sam et. al. "Chipping Away at Censorship Firewalls with User-Generated Content" to appear in the Usenix Security Symposium 2010. Available at http://www.gtnoise.net/papers/2010/burnett:usenixsec2010.pdf. Accessed July 11, 2010.

[15]Jackson, WIlliam. "Security will not come naturally with IPv6," *Government Computer News*, September 17, 2009. Available at http://gcn.com/Articles/2009/09/17/IPv6-security.aspx. Accessed May 31, 2010.

[16]Oudot, Laurent. "[Full-disclosure] TEHTRI-Security released 13 0days against web tools used by evil attackers." Available at http://archives.neohapsis.com/archives/fulldisclosure/2010-06/0423.html. Accessed July 1, 2010.

5.2 Preemptive Covert Operations

"Hack back," as the name implies, is a reactive strategy which is implemented upon detection of an attack; as with forensic-based methods, this may be too late. A further step in active defense is to mount preemptive covert operations against sites that are suspected to be planning or preparing attacks. These operations would focus on exploitation and intelligence-gathering, but could rapidly shift to suppression if an attack were detected. Awareness on the part of potential attackers that such "cyber patrolling" was being conducted would in and of itself act as a deterrent, since they would be faced with additional security tasks as well as the uncertainty as to the degree to which they may themselves have been penetrated and placed at risk.

5.3 Obstacles to Alternative Methods

There are no significant technical obstacles to the use of the techniques described above; they both are based on known and validated technology used by attackers themselves. Any obstacle that would potentially be raised would be of a legal and policy nature. One unresolved legal issue is the question of ownership of a subverted machine, such as an element of a botnet. One the one hand, it can be argued that the machine's Internet presence inherits the rights of the owner of the physical hardware. On the other hand, it can be argued that the machine's Internet presence is analogous to that of a vehicle which has been stolen and is being used in the commission of a crime, and which is then fair game for capture or suppression.

6 HUMAN ATTRIBUTION

The problem of converting technical attribution to human attribution is as old as crime and punishment. Barriers to it consist of all the known methods of evasion, deception, and denial. Consider, for example, a registered firearm which is linked by ballistics evidence to a crime. The person to whom the firearm is registered can, when confronted with this, simply claim that the gun was stolen. Similarly, a person who is associated with a machine known to be the source of malicious packets can simply claim that the packets were sent by malicious functionality that was installed without the person's knowledge. Given the power and deviousness of current techniques for doing so,[17] such claims are increasingly plausible.

Cyber forensics, as described above, are considerably less compelling than a ballistics test on a firearm. Whatever information is developed in the course of determining technical attribution must be combined with other information and analyzed in the context of the traditional investigatory triad of motive, means, and opportunity. The Internet adds further difficulties in that its transnational reach may, and in the criminal domain historically has, meant that investigations encounter severe jurisdictional constraints.

One technique that combines technical and human attribution, and that is limited in applicability but can yield valuable information, involves administrators establishing the means to capture and replay the real-time actions of an intruder who is conducting a reconnaissance exercise. Analysis of keystroke intervals, misspelling of command names, time of day and duration of intrusion and similar data can provide hints about the number, native language, and technical background of the group performing the intrusion. If multiple intrusions are detected, this form of analysis can also provide hints as to whether they represent a coordinated effort or are disjoint.

Whatever the technique, it is by no means certain that the benefits of effective human attribution would outweigh the adverse effects. Anonymity has been held to be an essential aspect of the effective

[17]Zetter, Kim. "Google Hack Attack Was Ultra Sophisticated, New Details Show." *Wired News*, Jan. 14, 2010. Available at http://www.wired.com/threatlevel/2010/01/operation-aurora/. Accessed May 31, 2010.

exercise of free speech.[18] A recent proposal by a member of the Obama administration for a voluntary system of Internet credentials has met with mixed response.[19] Besides the objections on the basis of personal privacy and freedom, there are serious questions of effectiveness and the degree to which any such system could resist techniques for identity theft. The latter concern is heightened by the phenomenon that any standardized and widespread security system becomes, by its ubiquitous nature, worth the expenditure of significant effort by hostile parties to defeat.

7 PERFECT TECHNICAL ATTRIBUTION AND DETERRENCE

7.1 Deterrence in General

The psychological theory of deterrence has been extensively studied, both for international relations and criminal justice.[20] The consensus is that individuals and organizations are deterred from aggressive action by two factors: first, that retribution is likely and will be unacceptably severe, and second, that risk of failure is too high. Before considering the role attribution may or may not play in deterrence, it is worthwhile to consider two factors which act to deter large-scale disruptive cyberattacks from any source.

The first of these is the risk of an unintended consequence that the initiator, or allies of the initiator, are harmed by the attack. For example, a crippling attack on the financial system of one nation by a transnational terrorist group carries the risk that the entire global economy may be adversely affected, which may in turn deplete the wealth of nations or individuals who are supporting that terrorist group.

The second factor is that a disruptive cyberattack is very unlikely to resemble a kinetic attack like a truck bomb, which achieves a near-instantaneous transition from normality to destruction. All but the most improbable scenarios for disruptive cyberattacks have them unfolding over time, time during which system administrators will be taking action to stop the attack and mitigate its effects. Thus the event will resemble a melee more than it will an explosion. The degree of uncertainty associated with such a widespread, dynamic, and interactive event and the associated difficulties of predicting success are an inherent obstacle to a decision to proceed with an attack.

7.2 The Deterrent Effect of Perfect Technical Attribution

The discussion to follow postulates an Internet in which there exists perfect technical attribution, that is, every action can be traced back to a specific human and every element of hardware and software can be identified as to source.

Whether perfect attribution would serve to deter an attack naturally depends on the nature of the actor considering a large-scale, society-disrupting attack. Given the scale of such attacks, and the magnitude of their possible consequences, these actors are not likely to be criminals or even large-scale criminal enterprises motivated by financial gain but rather state or non-state actors driven by geopolitical motives of power and influence.

There are four cases of where perfect technical attribution could be a factor: state-mounted attacks, state-sponsored attacks using non-state actors, state-tolerated attacks using non-state actors, and attacks by non-state actors with no state involvement.

State-mounted attacks are those conducted by the armed forces or covert action agencies of a nation state. The complex issues association with such attacks, as viewed from a U.S. perspective, are extensively treated in a previous National Research Council report.[21] Perfect technical attribution, in this case,

[18]Solove, Daniel J. *The Future of Reputation: Gossip, Rumour and Privacy on the Internet*, Yale University Press, 2007.

[19]Markoff, John. "Taking the Mystery Out of Web Anonymity," New York Times, July 2, 2010.

[20]McGill, William L. "Defensive Dissuasion in Security Risk Management," *Proc. 2009 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 3516-3521.

[21]National Research Council, *Technology, Policy, Law, and Ethics Regarding U.S. Acquisition and Use of Cyberattack Capabilities*. National Academies Press, 2009.

adds little to deterrence. If a nation is preparing a cyberattack of a scale which is near or at the boundaries of an act of war, then that attack will most probably take place in an environment of international tension or imminent armed conflict. Attribution therefore becomes obvious, and the planners of the attack will have incorporated the chances and nature of retaliation in their calculus.

The involvement of non-state actors complicates the problem of assigning ultimate responsibility for the attack, and the existence of perfect technical attribution reduces that problem to one of determining a relationship between an identified non-state actor and some state. In the case of sponsorship, this determination can be made through traditional investigative and intelligence gathering techniques such as tracing financial transactions, interception of communications, and so forth. This process is complicated, and attribution obscured, by a common practice of clandestine services known as "false flag operations." In these, the service pretends to be that of some other state, in order to gain the sympathy and cooperation of a non-state actor. The elaborate steps, such as routing of financial transactions, location of meetings, etc., which are taken to convince the non-state actor of the false state affiliation can also mislead the target of the attack when it attempts to determine upon whom to retaliate. The principal deterrent effect of knowing "who did it" in this case, will be against the non-state actor. Since that individual or group is under state sponsorship, it is likely they will also be under state protection, and the deterrent effect will be minimal.

Further complications are introduced by state-tolerated attacks. These occur when the so-called "patriotic hackers" of a particular nation independently launch attacks whose nature and timing coincide with the interests of that nation. In this case there are no overt, detectable links between the national authorities and the attackers; rather, all the authorities need to do is refrain from taking action, possibly while protesting publicly that they are doing all they can to stop the assaults. There has been media speculation that the attacks from China on the United States,[22] attacks on Estonia,[23] and attacks on the nation of Georgia[24] were state-tolerated. Even a small scale-attack can escalate by having other hackers in other countries "pile on" as word of the attack spreads.[25] Later reports suggest that the Estonian attack was actually triggered by an Estonian citizen.[26]

Here the link between the actor and the state is even more tenuous than in the previous case, further reducing the chance that the state will suffer retaliation. Since retaliation against the actor or actors will, in general, require state cooperation, they also have little reason to be dissuaded from attacking, even if they are identified through the mechanism of perfect technical attribution.

Finally there is the case of a non-state actor that is not sponsored by, nor independently acting in the interests of, a particular state. This is the one instance where perfect technical attribution may act to deter: since an attack by such an actor readily fits into the criminal justice domain, the actor has no sponsor or protector, and in many jurisdictions punishment will be severe, perfect technical attribution and the associated fear of likely and unacceptable retribution will act as a deterrent.

8 SUMMARY

1. The Internet contains intrinsic features and extrinsic services which support anonymity and inhibit forensic attribution of cyberattacks.

---

[22]Delio, Michael. "Is This World Cyber War I?" *Wired News*, May 1, 2001.Available at http://www.wired.com/politics/law/news/2001/05/43443. Accessed May 31, 2010.

[23]Landler, Mark and Markoff, John. "Digital Fears Emerge After Data Siege in Estonia." *New York Times*, May 29, 2007. Available at http://www.nytimes.com/2007/05/29/technology/29estonia.html. Accessed May 31, 2010.

[24]Krebs, Bryan. "Russian Hacker Forums Fueled Georgia Cyber Attacks," *Washington Post*, Oct. 16, 2008. Available at http://voices.washingtonpost.com/securityfix/2008/10/report_russian_hacker_forums_f.html. Accessed May 31, 2010.

[25]Anon. "A cyber-riot" *The Economist*, May 10, 2007.

[26]Hruska, Joel. "Student behind DoS attack that rekindled bad Soviet memories." *ars technica*, January 24, 2008. Available at http://arstechnica.com/business/news/2008/01/student-behind-dos-attack-that-rekindled-bad-soviet-memories.ars. Accessed May 31, 2010.

2. Recent events in copyright enforcement may accelerate the technical evolution anonymizing services.

3. There are few if any plausible ways to overcome the barriers to forensic-based attribution imposed by these features and services.

4. It is too early to state with confidence that the move to IPv6 will have a significant effect on the ability to attribute cyberattacks.

5. In cases where forensic-based technical attribution is possible, it is most likely to be achieved in the reconnaissance phase of an attack.

6. Alternatives to forensic-based attribution include counterattack ("hack back") and sustained, aggressive covert intelligence gathering on potential attackers. The obstacles to these methods are primarily nontechnical.

7. Even if perfect technical attribution were achieved, it would have a significant deterrent effect in but a minority of cases where significant disruptive cyberattacks are contemplated by parties hostile to the United States.

8. Preemptive covert operations may have significant deterrent effect by raising uncertainty of success owing to the possibility that facilities controlling an attack may contain latent subversions.