ESSAY

Untangling Attribution

David D. Clark* and Susan Landau**

I. Introduction

In February 2010, former Director of the National Security Agency Mike McConnell wrote, "We need to develop an early-warning system to monitor cyberspace, identify intrusions and locate the source of attacks with a trail of evidence that can support diplomatic, military and legal options — and we must be able to do this in milliseconds. More specifically, we need to reengineer the Internet to make attribution, geolocation, intelligence analysis and impact assessment — who did it, from where, why and what was the result — more manageable."[1]

The Internet was not designed with the goal of deterrence in mind, and perhaps a future Internet should be designed differently. McConnell's statement is part of a recurring theme that a secure Internet must provide better *attribution* for actions occurring on the network. Although *attribution* generally means assigning a cause to an action, as used here *attribution* refers to identifying the agent responsible for the action (specifically, "determining

* David Clark, Senior Research Scientist, MIT, Cambridge MA 02139, ddc@csail.mit.edu.
Clark's effort on this work was funded by the Office of Naval Research under award number N00014-08-1-0898. Any opinions, findings, and conclusions or recommendations expressed in this Essay are those of the authors and do not necessarily reflect the views of the Office of Naval Research.
** Susan Landau, Fellow, Radcliffe Institute for Advanced Study, Harvard University, Cambridge, MA 02138, susan.landau@privacyink.org. An earlier version of this Essay appeared in COMM. ON DETERRING CYBERATTACKS, NAT'L RESEARCH COUNCIL, PROCEEDINGS OF A WORKSHOP ON DETERRING CYBERATTACKS: INFORMING STRATEGIES AND DEVELOPING OPTIONS FOR U.S. POLICY, 25–40 (2010), *available at* http://www.nap.edu/catalog/12997.html.

[1] Mike McConnell, *Mike McConnell on How to Win the Cyber-war We're Losing*, WASH. POST, Feb. 28, 2010, http://www.washingtonpost.com/wp-dyn/content/article/2010/02/25/AR2010022502493.html.

the identity or location of an attacker or an attacker's intermediary.)"[2] This links the word to the more general idea of *identity*, in its various meanings. Attribution is central to *deterrence*, the idea that one can dissuade attackers from acting through fear of some sort of retaliation. *Retaliation requires knowing with full certainty who the attackers are.* **In particular, there have been calls for a stronger form of personal identification that can be observed in the network.**[3] A technically nonsensical but nonetheless clear complaint might be: "Why don't packets have license plates?" This is called the *attribution problem.* There are many types of attribution, and different types are useful in different contexts. We believe that what has been described as the attribution problem is actually a number of problems rolled together. Attribution is certainly not one size fits all.

Attribution on the Internet can mean the owner of the machine (e.g., the Enron Corporation), the physical location of the machine (e.g., Houston, Estonia, China), or the individual who is actually responsible for the actions. The differences between these varied forms of attribution motivate this Essay. Our goal is to tease apart the attribution problems in order to determine under which circumstances which types of attribution would actually be useful.

In summary, we draw the following conclusions:

1. Network-level addresses (IP addresses) are more useful than is often thought as a starting point for attribution, in those cases where attribution is relevant.[4]

---

[2] DAVID A. WHEELER & GREGORY N. LARSEN, INST. FOR DEF. ANALYSIS, TECHNIQUES FOR CYBER ATTACK ATTRIBUTION ES-1 (2003), *available at* http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA468859.

[3] *See, e.g.*, STEWART A. BAKER, SKATING ON STILTS: WHY WE AREN'T STOPPING TOMORROW'S TERRORISM 231–32 (2010), *available at* http://media.hoover.org/sites/default/files/documents/Skating_on_Stilts_Big_Brothers_Revenge_223.pdf (describing the proposals for attribution put forward by the former assistant Secretary for Policy at the Department of Homeland Security); CSIS COMMISSION ON CYBERSECURITY FOR THE 44TH PRESIDENCY, SECURING CYBERSPACE FOR THE 44TH PRESIDENCY 62 (2008), *available at* http://csis.org/files/media/csis/pubs/081208_securingcyberspace_44.pdf ("Creating the ability to know reliably what person or device is sending a particular data stream in cyberspace must be part of an effective cybersecurity strategy.").

[4] *See, e.g.*, W. Earl Boerbert, *A Survey of Challenges in Attribution*, *in* COMM. ON DETERRING CYBERATTACKS, NAT'L RESEARCH COUNCIL, PROCEEDINGS OF A WORKSHOP ON DETERRING CYBERATTACKS: INFORMING STRATEGIES AND DEVELOPING OPTIONS FOR

2. Redesigning the Internet so that all actions can be robustly attributed to a person would not help to deter the sophisticated attacks we are seeing today. At the same time, such a change would raise numerous issues with respect to privacy, freedom of expression, and freedom of action, a trait of the current Internet valued by many including intelligence agencies.

3. The most challenging and complex attacks to deter are those we call multi-stage attacks, where the attacker infiltrates one computer to use as a platform to attack a second, and so on. These attacks, especially if they cross jurisdictional boundaries, raise technical and methodological barriers to attribution.

4. A prime problem for the research community is the issue of dealing with multi-stage attacks. This — rather than the issue of designing highly robust top-down identity schemes — is the problem that should be of central concern to network researchers.

To illustrate the utility of different sorts of attribution, we will use several examples of attacks. First we consider a distributed denial of service (DDoS) attack. As we discuss below, one aspect of dealing with DDoS attacks involves stopping or mitigating them as they occur. (This aspect may or may not be categorized as "deterrence," or instead just as good preparation.) To stop a DDoS attack, we want to shut off communication from the attacking machines, which would most obviously call for attribution at the level of an IP address. On the other hand, to bring the attacker — the bot-master — to justice requires a different type of attribution. We must find a person, not a machine. Unlike the information for halting the attack, this form of attribution is not needed in real time. Next we consider a phishing attack, which attempts to extract information back from the recipient, so the attempted exploitation must include an IP address to which information is returned. The attribution question then becomes whether that address can effectively be translated into a higher-level identity (such as a person). Attribution in the cases of information theft

---

U.S. POLICY 41–52 (2010), *available at* http://www.nap.edu/catalog/12997.html (focusing on sophisticated attacks from state-sponsored agencies and concluding that attribution would not be a useful tool in those situations). For simpler and less sophisticated events, where one computer engages another directly, attribution may be a useful tool and we discuss the utility of IP addresses as a starting point for attribution in these cases.

can be easy (relatively speaking) if the information is used in criminal ways (e.g., to generate false identities and open fake accounts), but extremely hard if the stolen data, such as flight plans for U.S. military equipment, disappears into another nation-state's military planning apparatus.

We start by putting attribution in the context of Internet communications and then move to examining different kinds of cyberexploitations and the role attribution plays in these. We follow by considering attribution from four vantage points: type of identity, timing of attribution (before, during, and after an event), type of investigator, and jurisdiction. By considering both what information is available (through types of identity and timing of attribution) and what type of investigation is being done (type of investigator and particulars of jurisdiction), we are better able to discern what the real needs are for attribution.

II. Brief Introduction to Internet Communications

In common parlance, all parts of the Internet are often rolled together into a single phenomenon called "the Internet." Calls for better security are often framed in this simple way, but it is important to start with a more detailed model of the Internet's structure.

To its designers, the term "Internet" is reserved for the general platform that transports data from source to destination, in contrast to the various applications (email, the Web, games, voice, etc.), which are described as operating "on" or "over" the Internet. The data transport service of the Internet is based on *packets* — small units of data prefixed with delivery instructions. The analogy often used to describe a packet is an envelope, with an address on the outside and data on the inside. A better analogy might be a postcard, since unless the data is encrypted it too is visible as the packet is moved across the Internet.

The Internet is made up of a mesh of specialized computers called *routers*, and packets carry a destination address that is examined by each router in turn in order to select the next router to which to forward the packet. The format of the addresses found in packets is defined as part of the core Internet Protocol (IP), and they are usually referred to as IP addresses. Packets also carry a *source IP address*, which indicates where the packet came from (somewhat like the return address on a letter or postcard). This address thus provides a form of attribution for the packet. Since the

routers do not use the source address as they forward a packet, much has been made of the fact that the source address can be forged or falsified by the sender. For a variety of reasons, it is not always easy for a router to verify a source address, even if it tries.[5] However, since the source address in a packet is used by the recipient of the packet to send a reply, if the initial sender is attempting to do more than send a flood of one-way packets, the source address of the packet has to be valid for the reply to arrive back. For this reason, the source address found in packets often provides a valid form of source attribution.

Above the packet service of the Internet we find the rich space of applications — applications that run "over" the packet service. At this level, some applications employ very robust means for each end to identify the other. When a customer connects to a bank, for example, the bank wants to be very sure that the customer has been correctly identified. The customer similarly wants to be sure that the bank is actually the bank, and not a falsified web site pretending to be the bank. Encrypted connections from browser to bank,[6] certificate hierarchies, passwords, and the like are used to achieve a level of mutual identification that is as trustworthy as is practical.

There are two important points to note about these application-level identity mechanisms. First, the strength of the identification mechanism is up to the application. Some applications such as banking require robust mutual identity. Other sites need robust identity, but rely on third parties to do the vetting, e.g., credit card companies do so for online merchants. Some sites, such as those that offer information on illness and medical options, are at pains not to gather identifying information, because they believe that offering their users private and anonymous access will encourage them to make frank enquiries.

Second, these schemes do not involve the packets. An Internet engineer would say that these schemes do not involve the Internet at all, but only the services that run on top of it. Certainly, some of these identity schemes involve third parties, such as credit card companies or merchant

---

[5] One recent experiment concluded that nearly a third of Internet customers could spoof their source IP address without detection. ROBERT BEVERLY ET AL., UNDERSTANDING THE EFFICACY OF DEPLOYED INTERNET SOURCE ADDRESS VALIDATION FILTERING 1 (2009), *available at* http://www.caida.org/publications/papers/2009/imc_spoofer/imc_spoofer.pdf.

[6] The relevant protocols go by the acronyms of Secure Sockets Layer (SSL) and Transport Layer Security (TLS).

certification services. But these, too, are "on top of" the Internet, and not "in" the Internet.

In contrast to these two forms of identity mechanisms — IP addresses and application-level exchange of identity credentials, the "license plates on packets" approach would imply some mandatory and robust form of personal level identifier associated with packets (independent of applications) that could be recorded and used by observers in the network. This packet-level personal identifier, which might be proposed in the future for the Internet, is one focus of our concern.

III. Classes of Attacks

It has become standard to call anything from a piece of spam to a carefully designed intrusion and exfiltration of multiple files an "attack." However, lumping together such a wide range of events does not help us understand the issues that arise; it is valuable to clarify terminology. As a 2009 National Research Council report on cyberattacks delineated, some attacks are really *exploitations*. *Cyberattacks* and *cyberexploitations* are similar in that they both rely on the existence of a vulnerability, access to exploit it, and software to accomplish the task,[7] but cyberattacks are directed to disrupting or destroying the host (or some attached cyber or physical system), while cyberexploitations are directed towards gaining information. Indeed a cyberexploitation may cause no explicit disruption or destruction at all. We will use that distinction. Attacks and exploitations run the gamut from the very public to the very hidden, and we will examine cyberattacks/cyberexploitations along that axis.

A. Bot-Net Based Attacks

Distributed denial of service (DDoS) attacks, in which a large number of machines from all over the Internet attack a site or a small set of sites, have the goal of disrupting service by overloading a server or a link. They have a unique character: visible and intrusive. DDoS attacks are designed to be detected. The attack is done by first penetrating and subverting a large stock of attack machines, forming them into what is called a "bot-net." A DDoS attack is thus a multi-step activity, first building

---

[7] WILLIAM A. OWENS ET AL., TECHNOLOGY, POLICY, LAW, AND ETHICS REGARDING U.S. ACQUISITION AND USE OF CYBERATTACK CAPABILITIES 81 (2009).

the bot-net, then instructing the subverted machines to launch some sort of simultaneous attack on the target system. This step of the attack may be the sending of floods of packets or just overloading the server with apparently legitimate requests.

Before the attack, it may be possible to take active steps to reduce its potency. There are at least two approaches to degrading the attack's strength — making it harder to penetrate and keep control of a machine and identifying machines that are apparently infected, so they can be isolated if they participate in an attack. Machines that are seen as likely ultimate targets for DDoS attack can also prepare themselves by replicating their content on distributed servers, so that an attack must diffuse itself across multiple machines.[8]

During an attack, the relevant mitigation techniques involve turning off traffic from attacking hosts or discarding the traffic before it reaches the point of overload. This response requires knowing the identity of the attacking machines to identify the traffic. Note that it is not necessary to know all of the machines, just enough to reduce the attack to manageable proportions. And depending on what steps are taken to block traffic from the attacking machines, there may be minimal harm from the occasional misidentification of an attacker.[9]

After the fact, DDoS attacks represent a challenge for the objective of retribution. The attacker (the so-called bot-master or the client who has rented the bot-net from the bot-master) has usually taken care to be several degrees removed from the machines doing the actual attack. Tracing back through the attacking machines to find the responsible attacker may involve crossing jurisdictional boundaries, which adds complexity and delay. If the actual attack involved falsified source addresses, such trace-back may be very difficult or even impossible. However, the range of attacks that can be

---

[8] For example, a content provider might choose to outsource the hosting of its content to a Content Delivery Network (CDN). A leading provider of CDN service, Akamai, specifically claims that its infrastructure is massive enough that DDoS attacks will be ineffective against it. *See* AKAMAI SECURITY CAPABILITIES: PROTECTING YOUR ONLINE CHANNELS AND WEB APPLICATIONS 6–7 (2010).

[9] For example, if the mitigation technique involved blocking traffic coming from a source for a few minutes, then if an innocent machine were misidentified as part of the attack, the only consequence would be that the user of that machine could not reach the web site for that short time. That sort of failure can occur for many reasons and might well be the outcome that the user perceived in any case while the target machine was under attack.

executed without a two-way exchange of packets is very limited, and for many attacks today, the source address is not forged.[10] Because of these factors, there is a question as to whether after-the-fact-retribution is a useful part of dealing with bot-net-based DDoS attacks.

Bot-nets are also used to send bulk unsolicited email — spam. From an attribution perspective, this application is different from DDoS attacks. When bot-nets are used for sending spam, spam provides trace-back. Because merchants have to identify themselves in order to be paid, some attribution is possible. Spammers' protection comes not from anonymity, but from jurisdictional distance or legal ambiguity.

B. Identity Theft

The term "identity theft" has received much attention in the press recently, but it is worth separating the different activities that are sometimes lumped together under a single term. The Identity Theft and Assumption Deterrence Act of 1998[11] criminalized identity theft, which the Federal Trade Commission describes as "someone us[ing] your personally identifying information, like your name, Social Security number, or credit card number, without your permission, to commit fraud or other crimes."[12] Under this definition, up to nine million Americans suffer identity theft annually.[13]

This broad definition encompasses everything from the theft of a single credit-card number or misuse of a single account to a full-scale impersonation of an identity (involving the establishment of new credit accounts or identity documents in a person's name). The former constitutes the majority of identity theft. In 2006, for example, according to an FTC

---

[10] This statement does not imply that forged source addresses are never seen in current attacks. For example, some attacks are based on the use of the Domain Name System (DNS) as a vector, and those attacks are one-way and involve falsified source addresses. By sending a query to a DNS server with the source address of the machine to be attacked, the server will reply with a packet sent to that machine. *See, e.g.*, Daniel Weseman, *DNS Queries*, INTERNET STORM CENTER, http://isc.sans.edu/diary.html?storyid=5713 (last visited Feb. 18, 2011).

[11] Pub. L. No. 105-318, 112 Stat. 3007 (1998).

[12] *About Identity Theft*, FEDERAL TRADE COMMISSION, http://www.ftc.gov/bcp/edu/microsites/idtheft/consumers/about-identity-theft.html (last visited April 13, 2010).

[13] *Id.*

report, 6.5 million Americans suffered theft of their credit or account information, while 1.8 million had their identity information used to establish fraudulent accounts,[14] a ratio of about three-and-a-half to one. Thus the nine million number somewhat overstates the number of people subjected to full impersonation. The serious case of identity theft, in which new documents are established in someone else's name, happens about two million times a year in the United States.

Identity theft is an interesting crime for a number of reasons. It is a multi-step crime — the identity in question must be stolen and then exploited. The theft can occur in many ways. It may involve infiltration of a computer and installation of spyware that captures identifiers and passwords used for application-level authentication or the penetration of a merchant server and the theft of billing records. Such information may then be used by the original thief or sold to other criminals. Next, the identity must be exploited. If the exploit is on the Internet, this generally involves the use of the stolen credentials to mislead some sort of application-level authentication scheme, e.g., logging in as the user to lay a false attribution trail. Perhaps as a final step, some sort of money-laundering scheme is required to convert the exploit into money that is useful to the criminal.

Early Internet-based identity theft used "phishing," an attack in which a user is tricked into going to a web site that imitates a legitimate one (e.g., a bank) and typing in his name and password. Phishing attacks surfaced in 1996,[15] and by 2005, there were reports of as many as 250,000 phishing attempts being made daily against just one financial institution.[16] More lucrative than attempts at obtaining records about single individuals are efforts that download identity information about many individuals at once and then use that information to commit crimes.

One such incident involved a group from Russia and Estonia that, with the help of an insider, broke into a server at RBSWorldPay, an Atlanta-based card-processing company. Taking information on customer accounts

---

[14] FEDERAL TRADE COMMISSION, 2006 IDENTITY THEFT SURVEY REPORT 4 (2007), *available at* http://www.ftc.gov/os/2007/11/SynovateFinalReportIDTheft2006.pdf.
[15] GUNTER OLLMANN, THE PHISHING GUIDE: UNDERSTANDING AND PREVENTING PHISHING ATTACKS 4 (2004), *available at* http://www.windowsecurity.com/uplarticle/privacy/NISR-WP-Phishing.pdf.
[16] Christopher Abad, *The Economy of Phishing: A Survey of the Operations of the Phishing Market*, 10 FIRST MONDAY, no. 9, Sept. 2005, *available at* http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1272/1192.

— the card numbers and associated PINs and decrypting the protected information — the thieves created counterfeit debit cards, raised withdrawal limits on these accounts, and hired people for the day who withdrew 9 million dollars from 21,000 ATMs in 49 cities.[17] Another attack involved Heartland Payment Services, a major processor of credit-card and debit-card transactions. Heartland's systems were penetrated, and unencrypted data in transit between merchant point-of-sale devices and Heartland was sniffed (that is, read by the unauthorized software that had penetrated the network). The data collected included account numbers, expiration dates, and sometimes the account holder's name;[18] allegedly over 130 million accounts were compromised.[19]

The fact that internal bank and credit-card account records can now be accessed over the network has made theft of such records much easier. The pattern such as was employed in the RBSWorldPay case, in which a single insider transferred sensitive personal data to accomplices overseas, appears to be increasing in frequency.[20]

C. Data Exfiltration and Espionage

Foreign military and industrial espionage have long been problems for the United States. Prior to the ubiquitous use of the network in modern enterprises, such espionage required people-in-place to make contacts at target facilities, receive the stolen information, etc. Moles needed to be in place for years before they had access to desired information. Such an enterprise was an expensive and time-consuming proposition. For example, in order to acquire Western technical expertise, hundreds of Soviet case officers were involved in Soviet-U.S. collaborative working groups in

---

[17] Press Release, U.S. Dep't of Justice, Office of Public Affairs, Alleged International Hacking Ring Caught in $9 Million Fraud (Nov. 10, 2009), *available at* http://www.justice.gov/opa/pr/2009/November/09-crm-1212.html.

[18] Kevin Poulsen, *Card Processor Admits to Large Data Breach*, WIRED (Jan. 20, 2009, 12:40 PM), http://www.wired.com/threatlevel/2009/01/card-processor.

[19] Press Release, U.S. Dep't of Justice, Office of Public Affairs, Alleged International Hacker Indicted for Massive Attack on U.S. Retail and Banking Networks (Aug. 17, 2009), *available at* http://www.justice.gov/opa/pr/2009/August/09-crm-810.html.

[20] Dan Schutzer, *Research Challenges for Fighting Insider Threat in the Financial Services Industry, in* INSIDER ATTACK AND CYBER SECURITY: BEYOND THE HACKER 215 (Salvatore J. Stolfo et al. eds., 2008).

agriculture, civil aviation, nuclear energy, oceanography, computers, and the environment.[21]

The Internet has greatly simplified this process. Information that was once clearly inside a large enterprise may now be relatively easily accessible to people on the outside. Instead of all the work devoted to developing people-in-place, competitors, whether corporate or foreign governments, have discovered that the theft of secrets can be done over the network. Developing contacts, planting moles, and touring U.S. factories and development sites are efforts much less needed than once they were.

The first public reports of massive network-based data exfiltration surfaced in 2005. *Time* magazine reported a 2004 exploit in which U.S. military computers at four sites — Fort Huachuca, Arizona, Arlington, Virginia, San Diego, California, and Huntsville, Alabama — were, in a matter of six-and-a-half hours, scanned, and large numbers of sensitive files were taken. These materials were then apparently shipped to Taiwan, South Korea, and Hong Kong, and from there, to mainland China.[22] Since then numerous reports have surfaced of similar cyberexploitations, with the attempted intrusion method growing increasingly sophisticated over time.[23] The highly publicized intrusion into Google in 2009-2010 apparently followed this pattern.[24]

Attacks of this sort are stealthy and often of small scale. Frequently they are individually tailored. Their preparation may involve taking over insecure intermediate machines, but only in small quantities, and perhaps only those highly suited to the task. These machines are used to transit the stolen information and hide its ultimate destination. The first step in the theft is to carefully scope out the target, learning where the files of interest

---

[21] Matthew French, *Tech Sabotage During the Cold War*, FEDERAL COMPUTER WEEK, Apr. 26, 2004, http://fcw.com/articles/2004/04/26/tech-sabotage-during-the-cold-war.aspx.
[22] Nathan Thornburgh, *The Invasion of the Chinese Cyberspies*, TIME, Aug. 29, 2005, http://www.time.com/time/magazine/article/0,9171,1098961,00.html,
[23] THE U.S.-CHINA ECONOMIC AND SECURITY REVIEW COMMISSION, CAPABILITY OF THE PEOPLE'S REPUBLIC OF CHINA TO CONDUCT CYBER WARFARE AND CYBER NETWORK EXPLOITATION (2009), *available at* http://www.uscc.gov/researchpapers/2009/NorthropGrumman_PRC_Cyber_Paper_FINAL_Approved%20Report_16Oct2009.pdf.
[24] *See* John Markoff & David Barboza, *Inquiry Is Said to Link Attack on Google to Chinese Schools*, N.Y. TIMES, Feb. 19, 2010, http://query.nytimes.com/gst/fullpage.html?res=9505E7DA1730F93AA25751C0A9669D8B63.

are, and then, once target material has been located, to quickly pack and exfiltrate them, often in a matter of hours. The downloading may involve intermediate machines — "dead drops" — perhaps in South Korea, Taiwan, or Hong Kong, before the files are downloaded to their final destination (perhaps southern China). The multi-stage nature of the attack helps confound definitive knowledge of the ultimate destination of the files.

Investigation of such theft is very difficult. To trace back across the network to the perpetrator may involve several stages through multiple machines in different jurisdictions. However, the data being stolen must follow some path back to the perpetrator, which raises the possibility of tracking. Possession of the stolen information may or may not be useful as evidence, depending on the sort of retribution contemplated. From a national security perspective, these type of cases are the most important to deter. They are also the ones least likely to be solved solely through technical means.

IV. Cascades of Attribution and Multi-Stage Attacks

Many attacks and exploits are *multi-stage* in character: for example, A penetrates computer B to use as a platform for penetrating C, which is then used to attack D. Deterrence means focusing on computer A. It does not do much good to ask what person or actor owns machines B and C — they were just penetrated in passing. Following the chain of attribution backwards toward A involves identifying the IP addresses that lead back from D to C to B to A. If that trail can be followed, then the investigator can attempt to learn what can be discovered about A.

It is important to note both the limits of mechanisms for attribution and the intentional complexity of the various attacks and exploits, which have been crafted precisely to confound attribution. Looking at our earlier examples, we see patterns that are both *multi-step* and *multi-stage*. For example, a DDoS attack has a first step in which the array of attack machines (the bot-net) is assembled. This step will be taken in a multi-stage way, with the machines, as they fall prey to the initial event that infiltrates them, reporting back to some intermediate control computer that itself may have been first infiltrated and corrupted. Then in the step where the machines launch the attack, the instructions describing the attack will have been preloaded, and perhaps launched using a timer or a signal sent

through some complex signaling channel (e.g. a message to a chat channel), so that the controller is far away by the time that the attack is evident.

Of course, the multi-stage pattern is not unique to attacks and malicious behavior. Linking services together on multiple machines, such that A asks B to carry out some action, and B invokes C as part of the task, is the general idea behind composable services such as Web 2.0. In situations like this, A and B might exchange identity credentials, B and C might also do so, but C would not know whom A is.[25] B is providing a service to its clients (e.g., A), and uses C as part of this service. Under normal circumstances, B would take on the responsibility of ensuring that the clients (e.g. A) are not undertaking unsuitable objectives when they invoke the service. In case of a bad event (consider the analog of a multi-car rear-end collision), C complains to B, and B complains to A.

When the multi-stage activity is malicious, of course, the issue is that the intermediate machine has been infiltrated and corrupted, so the machine is not acting in a responsible way or in ways that reflect the wishes if its owner/operator. The human operator of B may be seen at the origin of the attack, but is just a victim of a security flaw in his machine.

One of the conclusions of this Essay is that multi-stage attacks must be a focus of attention when considering attribution and deterrence. First, many attacks fit in this category, including sophisticated and crafty attacks designed to avoid attribution. Assigning blame to such attacks is very challenging and difficult. Second, when computers are penetrated by an attacker to use as a platform for a further attack, that penetration usually bypasses any sort of end-to-end exchange of application-level credentials. **Therefore, the only kind of attribution that can possibly be applied here is at the level of IP addresses. Personal-level attribution will not be a useful tool in tracing attribution or assigning blame, and dealing with these sorts of attacks does not provide a justification for requiring network-based, personal-level identification.**

---

[25] One legitimate example of this occurs in federated identity management systems: the Identity Provider knows that Service Provider A and Service Provider B (for example, a hotel and a car-rental agency) are both providing services for the same customer, but through the judicious use of pseudonyms, no one else, including the two service providers, can determine that fact.

While multi-stage attacks represent a serious challenge, we urge the research community to consider what might be done to improve the options for tracking back to an ultimate source. Any solution or improvement that might be found will certainly not be purely technical, but will be a mix of technical and policy tools. For example, one might imagine every user of the Internet being urged to keep a log of incoming and outgoing connections. To avoid concerns about privacy, this log could be maintained under the control of the user himself — given today's technology, the sort of device called a "home router" could keep such a log with minimal additional cost for storage. But such a log would only be useful in a context where there are regulations as to when data could be requested from this log, by whom, etc. And of course, the user might have failed to maintain such a log. In such a case, the "punishment" might be that the ISP serving that user is required to log the user's traffic — the cost for failing to self-protect is a loss of privacy.

This idea may not be suitable — we offer it only as an example to illustrate how technology and policy will have to be combined as part of any solution, and also to illustrate that jurisdictional issues (and variation of regulation across jurisdictions) will be central in dealing with these sorts of attacks.

V. Four Different Aspects of Attribution

As the discussion above points out, different types of cyberattacks and cyberexploitations raise different options for prevention and deterrence. We have found it useful to think about attribution from different vantage points:

- Types: If users are expected to be identified in some way, what is the source of that identity, and what can we conclude about the utility of different sorts of identity?

- Timing: What are the different roles of attribution before, during, and after an event?

- Investigators: How might different parties exploit attribution as a part of deterrence?

- Jurisdiction: What are the variations that we can expect across different jurisdictions, and how might this influence our choices in mechanism design?

A. Types of Attribution

An IP address in a packet identifies an attachment point on the Internet. Roughly, by analogy to a street address, it indicates a location, but not who lives there. In many cases, of course, an address (both physical and Internet) can be linked to a person, or at least a family. Since residential Internet service is almost always provided by commercial Internet Service Providers (ISPs), they have billing information for all of their customers. If they choose to maintain a database that links billing information to the Internet addresses they give out to specific customers, they can trace back from address to personal identity. In the United States, the organizations that work to deter copyright infringement have had laws passed allowing them to obtain a subpoena for such information from ISPs. But unless this connection has been made, Internet addresses have meaning only at the level of a network endpoint, which usually maps to one or a small cluster of machines.[26] Indeed, in many cases, an IP address cannot be identified with a particular machine because the machine has been on the network for a quite short period of time, such as in an airport lounge, hotel lobby, or coffeehouse.[27]

In many application-level identity schemes such as the banking example above, identity has meaning at the level of an individual. The bank may keep track of Internet addresses as supplemental information to be used in case of abuse, but the design of their identity system is intended to tie directly to an individual as the accountable agent, not a machine. The IP address is not used as part of establishing that identity.

A related kind of individual identity is the *pseudonym*. The idea of a pseudonym, as the term is usually used, is an identity that links to a specific individual, without revealing who that individual is. A pseudonym system

---

[26] Many homes have a device called a "home router," which allows a small number of computers in the home to share one network connection. As the Internet is currently used, all these machines share one Internet address, so starting with that address there is no way to distinguish among those different machines. At a larger scale, an ISP (or a country) might use this same sort of technology to map a large number of machines to one address, making this sort of attribution even less effective.

[27] *See supra* text accompanying note 8.

should have two goals. First, the pseudonym should not be easily linked to an actual person — the goal is freedom from attribution. Second, the pseudonym should not be easily stolen and co-opted by another individual — the speaker, although anonymous, should have the exclusive use of that identity. Encryption schemes can be used in various ways to achieve this sort of functionality, which is a sort of "anti-attribution."

To fully protect pseudonymous speech and other types of anonymous activities, it is necessary to complement application-level "anti-attribution" mechanisms with tools to mask IP-level machine-based identity, since that can often be linked to human-level identity with some effort, as discussed above. Tools such as Tor[28] are used to give IP-level anonymity to communications; they are employed by activists and dissidents, journalists, the military and the intelligence community, and many others to mask with whom the communication is occurring. Law enforcement uses Tor to visit websites and chat rooms without leaving behind a tell-tale government IP address, while the military uses Tor to enable personnel "in place" to communicate with headquarters without revealing their true identity.[29]

When Internet communications occur without the use of traffic analysis anonymizers such as Tor, the source and destination addresses in packets can be seen by every router that forwards the packet, and by any other sort of monitoring device that is in the path from the sender to the receiver. As such, these sorts of identity indicators are fairly public. In contrast, if two end points exchange identity credentials between themselves over an encrypted connection, that exchange is private to those two end-

---

[28] Tor is a tool developed by the U.S. Naval Research Lab to permit anonymous (at the IP level) use of the Internet. *See* TOR, https://www.torproject.org/ (last visited Feb. 25, 2011).

[29] Tor allows anonymous communication and by its very nature does not reveal who is communicating with whom. Thus one cannot point to specific instances where the system was used to support military communications. However, various sources have claimed that Tor is in fact used by the military. *See, e.g.*, *Granting Anonymity*, N.Y TIMES MAGAZINE, Dec. 17, 2010, http://www.nytimes.com/2010/12/19/magazine/19FOB-Medium-t.html?_r=1&ref=virginiaheffernan. The fact that the project had its genesis in the Naval Research Laboratory and was funded for many years by the Defense Advanced Research Projects Agency is a clear indication of the value of the program to national security communications. Tor is available for public use because "anonymity loves company," that is, broad use of the system by outsiders hides the instances of national security and law enforcement communications. *See* Roger Dingledine & Nick Mathewson, *Anonymity Loves Company: Usability and the Network Effect*, *in* Security and Usability: Designing Systems that People Can Use, (Lorrie Cranor & Simson Garfinkel eds., 2005).

points.[30] Even if a third party, such as a credit-card company, is involved in the identity verification, that third party has been invoked with the knowledge and concurrence of the initial end-points. The knowledge of the identity is restricted to those parties.

An analogy to monitoring IP addresses in the network might be security cameras. A camera on a public street captures our public behavior and a likeness of our face. But, it does not reveal who we are unless that face can be linked to some other aspect of identity. In contrast, in various circumstances we have to identify ourselves to some other entity (show a driver's license, passport, credit card, etc.), but this transaction is specific to the circumstances at hand, and is normally not visible to a third-party observer. A security camera in a store provides an analogy to the logging of IP addresses by an endpoint. The images might be more easily linked to a customer transaction, and thus to other aspects of identity. But, the video captured by that camera is private to the store unless it chooses to reveal it (e.g., after a robbery) or it is demanded by an authorized third party (e.g., by a court order).

Using IP addresses as a starting point, one can try to derive forms of attribution other than at the level of the individual. IP addresses are usually allocated in blocks to Internet service providers (ISPs), corporations, universities, governments, and the like. Normally, the "owner" of a block of addresses is publicly recorded, so one can look up an address to see to whom it belongs. This can provide a starting point for investigation and subsequent fact-finding.

Another potential form of attribution is the "where" — geo-locating the end-point associated with the IP address on the face of the physical landscape. IP addresses are not allocated in a way that makes geo-location automatic — they are given out to actors that may have large geographic scope. Nonetheless, for many IP addresses, one can make a very accurate guess about where the end-point is located, since many networks have a hierarchical design to their physical connectivity, and map the addresses to

[30] The restriction of *encrypted* communication is critical here. If the observer is using technology called Deep Packet Inspection, or DPI, he can observe anything not encrypted, including identity credentials being exchanged end-to-end. Encryption does not hide everything; it is possible, for example, to determine the type of traffic (e.g., VoIP or video) even while the content itself is hidden.

the levels of the hierarchy. Several commercial services now exist that provide the function of mapping an IP address to an approximate location.[31] These services are designed to meet a number of customer needs, as their advertising suggests, including customization of Web content to different classes of customers and regulatory compliance. These services compete to provide accurate location information and advertise their precision in their marketing information. Various firms claim that 99-99.9% of IP addresses can be accurately localized to within a country, and that 90-96% can be accurately localized to within a state, city, or other similar region.[32] These services are used today by commercial Web content providers to localize their content to the presumed location of the user (e.g., to pick the right language), or in some cases, to block access to certain content based on the presumed locus (with respect to a jurisdiction), such as the blocking of Nazi memorabilia auctions to customers in France.[33] They are designed to work in real-time (as part of processing a Web query), and can provide a rich, if approximate, mapping from IP address to other attributes.

The issue with many of these tools is that since the mapping is approximate, there is some degree of "plausible deniability" to assertions of responsibility. There have been proposals to "harden" the linkage between IP address and other information. For example, several countries put forward a proposal to the International Telecommunications Union (ITU) that as part of the conversion of the Internet from IPv4 to IPv6,[34] addresses should be first allocated to states, which would then allocate them to the relevant private-sector actors. This would mean that the linkage from IP

---

[31] *See, e.g., MaxMind GeoIP Databases,* MAXMIND, http://www.maxmind.com/app/city (last visited Feb. 18, 2011); *Our Technology, NetAcuity® IP Intelligence: Paving the Way to Deeper Online Connections,* DIGITAL ELEMENT, http://www.digitalelement.com/our_technology/our_technology.html (last visited Feb. 18, 2011); *Services: We Make High-Volume Transaction Seamless*, QUOVA®: KNOW WHERE, http://www.quova.com/what/services/ (last visited Feb. 18, 2011).

[32] For example, the MaxMind service cited above states that it is "99.8% accurate on a country level, 90% accurate on a state level, 83% accurate for the US within a 25 mile radius." *Id.*

[33] For a discussion of the French litigation to block the sale of Nazi memorabilia in their country, see JACK GOLDSMITH & TIM WU, WHO CONTROLS THE INTERNET:: ILLUSIONS OF A BORDERLESS WORLDS 1–6 (2006).

[34] The addresses currently used in the Internet, called IPv4 addresses, are not sufficiently large to deal with the growing size of the Internet. To deal with all the devices that are anticipated to be attached in the future, a new and larger set of addresses, called IPv6 addresses, have been designed. A transition from IPv4 to IPv6 is now beginning.

address to jurisdiction would be robust,[35] and that it would be possible, for example, for the Chinese government to be certain where downloaded material, whether software stolen from U.S. companies or human-rights information from U.S organizations, was going.

Of course, the transition from IPv4 to IPv6 is only one of the changes that may occur to the Internet over the coming years. A more dramatic change might be the introduction of a virtualized network infrastructure, which would permit multiple simultaneous networks to co-exist, each with its own approach to attribution.[36] A future network that provides an information dissemination and retrieval service as part of its core function would imply some sort of binding between user and information that would be visible "in the network." We believe that our general conclusions will apply across a range of possible future network designs — the linkage between machine-level attribution and higher-level attribution (e.g. personal) will be a jurisdictional policy matter, not just a technical matter, and mechanisms for attribution must balance a range of policy objectives, not just focus on deterrence.

B. Timing

It turns out that timing — whether one is attempting to protect against a bad situation from developing, stopping an attack in its tracks, or investigating an exploitation after it has occurred — affects the methods one uses in handling the problem. Thus it is useful to consider attribution from these various vantage points.

---

[35] There is some disagreement as to whether the original proposal was for *some* or *all* IPv6 addresses to be allocated to countries. For a 2004 statement that makes clear that the proposal for only for *some* addresses to be allocated in this way. see H. ZHAO, INT'L TELECOMM. UNION, WORKING GROUP ON WORLD SUMMIT ON INFORMATION SOCIETY, ITU AND INTERNET GOVERNANCE 8–9 (2004) *available at* www.itu.int/ITU-T/tsb-director/itut-wsis/files/zhao-netgov02.doc.

[36] One way to understand virtualization is to continue the analogy to delivery of letters and postcards. Instead of using separate physical trucks for different delivery services — the postal service, UPS, Fedex, and so on — the various providers could decide to have one physical fleet of trucks that is "virtualized," in other words shared among all the providers. Each truck would follow only one physical route, but the different services might have different formats for addresses. Of course, to complete the "virtualization," not only would the space inside the truck be shared, but the truck would cleverly change the logo on the side as needed so it always had the correct branding to the customer. Each delivery company would have a "virtual truck" driving down the street.

1. Before the Fact: Prevention or Degradation

Actions taken before the attack are the ones most commonly associated with "computer security" — they involve good defenses for computers and the networks themselves, such as by downloading the latest patches or instituting good operating practices. None of these involve the need for attribution, but putting tools in place to implement good authentication and authorization are part of good security. For some classes of attacks, specifically DDoS events, it may be possible to degrade the viability of the bot-net or the potency of the attack by preventive actions that affect infected machines. In this respect, degradation of attacks can involve remote attribution.

2. During the Fact: Mitigation

During an attack/event, the main objective is to stop or mitigate the event. Secondarily, one may want to gather evidence to be used after the fact. What one can do during an attack depends on the nature of the attack, and different approaches to mitigation place different requirements on attribution for the attack. Different approaches will be needed to stop a DDoS attack and data exfiltration while it is happening.

3. After the Fact: Retribution

The traditional discussion of deterrence focuses on what would happen after the fact, when some sort of retribution would be exacted. For example, as discussed above, if the event is classed as a crime, this would trigger a police response. Primarily, police investigate crimes, identify the perpetrator, and gather the evidence for prosecution. Attribution is at the center of this role. Unless one can identify the perpetrator, retribution is hard to achieve. However, as we illustrated above in our examples of attacks, the actual situation is more complex in a computer-generated situation than this simple story might imply.

4. Ongoing: Attribution as a Part of Normal Activity

In fact, the "before the fact" phase above defines what should be the normal operating mode of the system. With good preparation, bad events might not occur. However, one should look at the role of identity and attribution in the ongoing operation of a system. The idea of authentication

is well understood. Several sorts of ongoing activities are made more trustworthy not by trying to prevent misbehavior in real time, but by demanding strong accountability. For example, access to medical records in an emergency room may best be controlled by allowing the access but requiring that the doctor making the request be thoroughly identified so the request can be logged.

C. Investigators

There are various sorts of deterrence that might be imagined; these have different implications for the needed quality and precision of the attribution. Different actors — police, intelligence services, and the military — will benefit from different sorts of attribution. In the case of attacks that are described as crimes, the usual sort of deterrence is judicial — arrest and prosecution — while in the case of cyberexploitation from military or national security sites, the deterrence may take diplomatic or retaliatory routes.

Judicial response would seem to call for attribution at the level of the individual, and of *forensic* quality — sufficient to bring into court. However, this model of attribution may be over-simplified. First, the most important role of attribution may be during the course of the investigation, when evidence is being gathered. Having a clue about attribution that is sufficient to guide an ongoing investigation may be critical. One FBI agent put it this way: "I could do packet attribution and let's say it gets me to a physical location. Maybe I get a search warrant and I get back. How I get there is important."[37]

After that point, forensic quality evidence matters. From the investigator's standpoint, "[What's] critically important is that you have evidence. Packet attribution is not beyond a reasonable doubt. The biggest thing in attribution is you're not looking for a computer; you're looking for a person."[38] Prosecutors look for certain kinds of evidence to bring before a jury. Evidence of on-line identity, however robust technically, may be less compelling than evidence gathered from carrying out search warrants and following the money. Packet-level attribution may aid an investigation, but our world still demands that the real evidence come from the physical world.

---

[37] Telephone interview by Susan Landau with senior FBI official (Dec. 14, 2009) (notes on file with Landau).
[38] *Id.*

National security investigators perform a different act than law enforcement investigators when sifting the evidence in a cyberexploitations or cyberattacks. They are seeking intelligence rather than producing court evidence. Where that evidence is produced — the jurisdiction — will play an important role in its veracity. A national security investigation cannot depend on packet-level attribution produced outside a trustworthy domain.

D. Jurisdiction

Different parts of the Internet operate within different jurisdictions: different countries, different legal systems, and (within these jurisdictions) both as public and as private-sector activities. Any discussion of attribution must consider jurisdictional issues.

1. Variation in Enforcement

Some regions may be lax in their enforcement of laws and uninterested in making the investigation of cyber-attack a high priority. This can be an issue in any attack, but becomes of particular importance in attacks that involve cascades of machines: machine A infiltrates machine B to attack machine C, and so on. If the jurisdiction within which B sits is not responsive, it becomes much harder to gather any evidence (which may be transient) that might link B to A. There is anecdotal evidence that attackers may "venue-shop" for regimes in which aggressive investigation is unlikely.

Evidence suggests that for single-stage events, so long as there are procedures in place within a jurisdiction, mapping from IP address to higher-level attribution is practical. For example, in the United States, the Recording Industry Association of America ("RIAA"), under the provisions of the Digital Millennium Copyright Act, regularly obtains information from ISPs about their customers hosting material covered by copyright for the purpose of bringing lawsuits.[39] The conclusion reached from this example should be the importance of jurisdiction in such a network investigation. To determine traffic origin requires investigating the machines traversed by the communications. If a jurisdiction permits such an investigation, then attribution — and perhaps deterrence — is possible. But if it does not, say because the jurisdiction does not view the activity as criminal, then tracing will not be possible.

---

[39] *See, e.g.*, Recording Indus. Ass'n of Am., Inc. v. Verizon Internet Serv., 351 F.2d 1229 (D.C. Cir. 2003).

This suggests that even if we were to push for a variant of the Internet that demanded very robust identity credentials to use the network, tracing would remain subject to barriers that would arise from variation in jurisdictions. Unless we imagine that all countries would agree to the election of a single, global identity authority, credentials would be issued by individual countries, where the quality of the process would be highly variable. In view of this, it is worth examining the issue of criminal versus national security investigations more closely.

2. Criminal Versus National Security Investigations

"Follow the money" is surprisingly useful as a guide to investigations. That adage might seem odd in investigating crimes that are purely virtual, but the fact is that almost all criminal activity (including child pornography) involves money. Thus, for example, although their initial theft was of bits, if the RBSWorldPay[40] criminals were to profit, in the end they needed to collect money from bank accounts. Even in child pornography cases, there are producers, organizers, users — and money.

A lack of laws against criminal activity on the Internet originally made prosecution of such activities difficult. Thus, for example, there were no charges brought against the Filipino developer of the 2000 ILOVEYOU virus; the Philippines only criminalized this activity three months after the release.[41] A combination of the development of national laws and much greater international cooperation has greatly improved the ability to track and prosecute clearly criminal Internet activities (e.g., identity theft, child pornography, and malware propagation). The key issue is what constitutes "clearly criminal." Economic espionage is not a crime in much of the world, and therefore other nations are unlikely to aid the United States in investigating or prosecuting such activities conducted against U.S. industries. That does not mean that investigation and consequences are not possible, only that they cannot follow the path of criminal prosecution the way as did, say, theft from RBS WorldPay.

If a nation-state is involved in data exfiltration, then the problem is a national security issue, not a law enforcement case. The level of proof of the attribution need not stand up in court. Indeed, the level of proof used to

---

[40] See *supra* text accompanying note 23.
[41] CIVIL CODE, Rep. Act 8792 (Phil.).

determine the attribution may never be made public even if the accusations of spying are. Intelligence agencies deal with certain forms of espionage, such as cyberexploitations of national research labs or defense contractors. Intelligence agencies do not usually try to bring spies into court — governments have their own ways of pushing back on attacks — but instead employ forms of tit-for-tat that require a degree of attribution, but perhaps only at the level of the state actor responsible. Diplomats can enter into a "shall we confront or cooperate" negotiation with their counterparts, using evidence that might not stand up in court but which is sufficiently compelling to underpin the negotiation.

Finally, if a cyberattack occurs as part of what is seen as "armed conflict," there may be some form of military response. This form of response is not usually directed at a specific person, but at a state or non-state group. The level of attribution that is required is thus to some larger aggregate, not the individual. To the extent that the initial manifestation of attribution is at the level of the IP address, the question that arises is how, and with what precision, this can be associated with some collective actor. To the military, attribution at the level of an individual is not useful. Of course, there may be times when it is in a nation's interest not to publicly attribute hostile acts to other nations — even when they are sure which nation is the hostile actor. Thus packet-level attribution would actually be against a state's interest.

VI. Summarizing the Value of Attribution

While there are probably many specific identity/attribution schemes, they seem to fall into general categories: the *machine,* the *person*, and the *aggregate identity***,** such as a state actor. The term *principal* is often used to describe the person or other entity that is ultimately accountable for some action.

Machines may have their own credentials and may store credentials for principals, but machines act only on behalf of some agent, and that agent (individual or collective) is the entity that must be identified and held accountable if effective deterrence is to occur. Thus machine attribution plays an important role in attribution, but is not of great value by itself if the goal is holding that agent accountable.

Under many circumstances, it is possible, with some effort, to link an IP address to a higher-level form of identity, whether an individual, a family (for residential broadband access), a corporation, or a state. Making this connection may be very difficult if the alleged attacker is in another jurisdiction. More importantly, attacks that involve cascades of machines challenge us to make the linkage back to the computer that belongs to the attacker that should be held accountable.

During an attack, when the goal is mitigation, it is not generally useful to identify the responsible person; what is needed is to deal with the machines that are the source of the attack. This sort of attribution is usually associated with IP addresses.

Retribution is not typically directed at a machine; after all, one does not usually arrest a machine. However, one could imagine various forms of active defense, in which a system under attack reaches out and somehow disables the attacking machine. This could be seen as a form of tit-for-tat retribution. It is probably illegal under U.S. law, but would represent an example of punishing a machine rather than a person. The practical issue here is that if the machine is an intermediary belonging to an innocent user, the degree of punishment (if it is allowed at all) must be carefully crafted to fit the crime. Mitigating these sorts of attacks is important, and various proposals will have to be considered, such as asking the ISP hosting an attacking machine to disconnect it from the net for a few minutes. Any such scheme must be designed in such a way that it itself cannot be subverted into a tool for originating an attack. One might force a machine to reboot to see if this disabled the attack code, but this again looks like a direct attack.[42]

A. What Attribution Can Deliver

Various different approaches are possible: machine-level attribution, application-level attribution based on credentials exchanged between end-points, and redesigning systems so the costs of an attack lie partially on the attackers. We consider each of these briefly.

---

[42] Current recommended practice for ISPs is for the ISP hosting the infested machine to verify that the machine appears to be part of a bot-net, then use its billing records to translate from machine to person, and send the person a letter.

1. Machine-Level (IP address) Attribution

Much has been made of the fact that source IP addresses can be forged. However, the only sort of attack where a forged IP address is effective is a DDoS attack, where the goal is just to flood the destination with useless traffic. Any more sophisticated exchange, for example in support of espionage, will necessitate a two-way exchange of information; this requires the use of valid source addresses. In a multi-step attack, the infiltration preparation of the intermediate machine requires meaningful communication; all but the last step will have valid source addresses.

2. Application-level Attribution

Especially if we were to redesign some protocols, the use of application-level attribution based on credentials exchanged among end-points is the approach that has the best balance of implications. First, the applications, having knowledge of the task, can pick the best tradeoff between strong accountability and the resulting protections and weaker (or no) accountability and its freedoms. A web site may want to allow access without demanding any identification, even though doing so weakens its access to retribution for attack. The site can compensate for this by limiting the consequences of attack — certainly there should be no confidential information on such a machine. DDoS may be the only real peril for such a machine, since defacement can usually be corrected quickly.

On the other hand, a machine storing highly confidential information should have no reason to permit any connection without strong identification of the other parties. An example of such would be a system handling process control for the power grid. While the best security would be to have the system not connected to the public Internet, for reasons of convenience, such connections do occur. In that case, one wants a defensive strategy that would include connection to the system only if strong forms of authentication are employed.

If a machine is attacked, we need a regime in which that machine can present evidence of attribution that it has gathered (both at the IP and application level), which it chooses to reveal because of the attack. Steps must be taken to prevent the end-point from falsifying this evidence; for example by means of some use of cryptography, or the use of trusted observers as witnesses. If this approach can be made to work, then the

revelation of each party's identity is under the control of the other parties, but no others. This seems like a nice balance of features.

3. Approaching Attribution Orthogonally

One might conclude from the above discussion that the goal of improved deterrence based on better attribution is hopeless. This conclusion is overly pessimistic. The correct conclusion we draw is that change to the Internet to add some sort of public, personal identity mechanism at the packet level is not useful and in fact counter-productive. Such identification would adversely impact privacy and would seriously impair many law enforcement and national security investigations. But one might imagine various sorts of clever "shifts in the playing field" that would make certain sorts of attribution easier to accomplish. Thus one could indirectly — orthogonally — approach the issue of attribution.

For example, would allocation of addresses to countries so that addresses could more easily and robustly be linked to a jurisdiction be a good idea? Such a change would have many implications,[43] and careful thought would be required to consider whether such a change would be in the best interest of nation-states, ISPs, content providers, Internet governance organizations, users and the other actors involved with the Internet.

Would it make sense to hold owners of intermediate machines in a multi-stage attack responsible to some (perhaps minor) degree for the resulting harm of the attack? This approach might heighten attention to better security of computers attached to the net and might lay the groundwork for a multi-stage trace-back system in which machines that allow themselves to be infiltrated become subject to third-party external surveillance as a consequence. To put it another way, poor system maintenance would result in a loss of privacy.

---

[43] Implications might include state control over who can have addresses and more state control over which firms can be ISPs, better localization of content in country-specific ways, selective jurisdiction-specific blocking of content, and reduction in the power that regional address allocation authorities have over the operation of ISPs within different countries.

4. Costs of Attribution

Few technical solutions have purely one-sided effects, and attribution is no exception to this general principle. Once a mechanism for attribution is put in place, we must expect that it will be used differently in different jurisdictions, according to the laws and customs of each country. In the United States, we may talk about deterrence as a goal to stop the breaking of our laws, while another country might use the same tools to repress dissidents. Better attribution tools could also be used to detect our intelligence services at work. Making one task easier makes these other related tasks easier, unless we take specific actions to separate classes of activity in a technical way. This sort of separation would imply the use of different forms of attribution in different circumstances; a consequence of this is that attribution tools should not be built into the core fabric of the Internet.

VII. Conclusions

Our fundamental conclusion is that "the attribution problem" is not actually a technical issue at all, but a policy concern with multiple solutions depending on the type of technical issue — e.g., DDoS attack, criminal activity, or data exfiltration — to be solved. Our conclusions are that, not surprisingly, solutions to the "attribution problem" lie outside the technical realm, and are instead in the space of law, regulation, multi-national negotiation, and economics.

A. Conclusion 1

The occasions when attribution at the level of an individual person is useful are very limited. Criminal retribution requires identifying a specific person and assigning blame, but the evidence that is finally brought into court is unlikely to be "forensic quality" computer-based identity, but rather other sorts of physical evidence found during the investigation. Clues about identity may be more important during the course of an investigation.

B. Conclusion 2

There is an important distinction between what we call private attribution (private to the end-points) and public or third-party attribution.

In application-level attribution as we described it, each end-point may take steps to know who the other parties to the communication are, but that knowledge is private to the communicating parties. In public or third-party attribution, an "observer in the middle" is given enough information that it can independently identify the communicating parties. In the current Internet, the only form of observer attribution is based on IP addresses. Where public attribution is useful, it will be at the level of the machine, not the person. The most obvious case is "during the fact" DDoS mitigation, where nodes in the network need to take action based on source and destination addresses.

We believe that public attribution beyond what is available today (that is, not based on the IP address, but on finer levels that would identify a user) would seldom be of value in the Internet, and would, at the same time, be a major threat to privacy and the right of private action. Such a change would be inimical to many values intrinsic to the United States, including rights protected by the First Amendment to read and write anonymously.[44] As a corollary, we note that there are two kinds of observers, trusted (by one of the end points) and untrusted (or unaffiliated, perhaps). If and when observer-based attribution is useful, it will often be a specific case where one of the end-points invokes a trusted observer to monitor what is being sent, perhaps as a witness, or because the end-point machine is not itself trusted.

C. Conclusion 3

Multi-stage attacks, which require tracing a chain of attribution across several machines, are a major issue in attribution today.

This problem can be confronted in a number of ways, including by making hosts more secure (a long-term effort) and by making it harder for an infested machine to launch a subsequent attack. If this problem could be resolved, it would eliminate many uncertainties in attribution. Since it is not now resolved, it imposes limits on the utility of attribution, no matter how it is structured. **Thus a prime problem for the research community is the issue of dealing with multi-stage attacks. This should be of**

---

[44] This right is not unqualified. For example, in Stanley v. Georgia, 394 U.S. 557 (1969), the Supreme Court ruled that states could not criminalize private possession of obscene materials. For a thoughtful discussion of First Amendment protection on the right to read anonymously, see Julie Cohen, *A Right to Read Anonymously: A Closer Look at "Copyright Management" in Cyberspace*, 28 CONN. L. REV. 981 (1996).

central attention to network researchers, rather than (for example) the problem of designing highly robust top-down identity schemes. Long term, we should look at what sorts of attribution would be of value if the multi-stage attack problem had been mitigated, as well as what is useful now.

Any attempts to deal with multi-stage attacks by tracing back the chain of machines involved will depend more on machine-level attribution at the intermediate steps, rather than personal-level attribution. Since the intermediate machines are normally being used without the permission (or knowledge) of their owners, knowing the identity of those owners is not very useful in trace-back. While one might imagine holding those owners accountable for some sort of secondary responsibility, the primary goal is to find the primary actor responsible for the attack, which involves following a chain of connections between machines.

D. Conclusion 4

We believe that pragmatically, the most important barriers to deterrence today are not poor technical tools for attribution, but issues that arise due to cross-jurisdictional attacks, especially multi-stage attacks. In other words, deterrence must be achieved through the governmental tools of state and not by engineering design.

Shifting the national security problem of attribution to its proper domain, namely from the tools of technology to the tools of state, means several changes in thinking about how tackle the problem. Rather than seeking solutions to the broad "attribution problem," networking researchers should move to considering the more narrowly focused problem of multi-stage attacks. Instead of seeking a purely technical fix, the U.S. government should move to diplomatic tools, including possibly treaties on cybercrime and cyberattack, to handle the multi-stage, multi-jurisdictional challenges of cyberexploitation and cyberattack. The efforts for top-down control of user identity and attribution, while appropriate and valid for critical-infrastructure domains such as those of energy, financial, and government services, have little role to play in the broader public network. Such efforts can be avoided, leading ultimately to better public safety, security, and privacy.