

Role and Challenges for Sufficient Cyber-Attack Attribution

**Authors:
Dr. Jeffrey Hunker,
Bob Hutchinson,
Jonathan Margulies**

January 28, 2008

Acknowledgments:

This work was supported under grant number 2003-TK-TX-0003 from the U.S. Department of Homeland Security, Science and Technology Directorate. Points of view in this document are those of the authors and do not necessarily represent the official position of the U.S. Department of Homeland Security or the Science and Technology. The I3P is managed by Dartmouth College.

Copyright© 2008. Trustees of Dartmouth College.

1	<u>INTRODUCTION</u>	5
1.1	A motivating example	5
2	<u>WHAT IS ATTRIBUTION?</u>	6
2.1	Attribution defined	6
2.2	Why, and when, attribution is important	7
2.3	The difficulty of attribution	7
2.3.1	Choices of Internet architecture	7
2.3.2	Administrative and policy issues	7
2.3.3	Limited business and economic support	8
2.3.4	Other impediments to attribution	8
2.4	What is adequate attribution?	8
2.4.1	Attribution can take many forms	9
2.4.2	Nature of the attack	9
2.4.3	Intended use or purpose for adequate attribution	9
2.5	When attribution is not desirable	12
2.6	A purely technical solution is impossible	13
3	<u>ATTRIBUTION POLICY</u>	14
3.1	Policy challenges to attribution	14
3.1.1	Incentives for Internet-wide solutions	14
3.1.2	Establishing trust across administrative or jurisdictional boundaries	14
3.1.3	Balancing privacy and attribution	14
3.1.4	Liability	15
3.1.5	Evidence preservation	15
3.1.6	Administrative context of attribution activities	15
3.2	Domestic legal authorities	16
3.2.1	Interception of communications	16
3.2.2	Cooperation among law enforcement and intelligence services	17
3.2.3	Limitations on intelligence activities	17
3.2.4	Cooperation of intermediate ISPs and systems administrators	17
3.3	International cooperation	18
4	<u>ATTRIBUTION TECHNOLOGY</u>	19
4.1	Technical challenges to attribution	19
4.1.1	Entry point anonymity	19
4.1.2	Retention capabilities of routers	19

4.1.3	Securing attribution mechanisms	19
4.2	Selected current technical approaches to attribution	19
4.2.1	Hash-based IP traceback	20
4.2.2	Network ingress filtering	21
4.2.3	ICMP return to sender	21
4.2.4	Overlay network for IP traceback	22
4.2.5	Probabilistic packet marking	23
4.2.6	Generating trace packets (iTrace)	23
4.2.7	Hack-back	23
4.2.8	Honeypots	24
4.2.9	Watermarking	25
5	<u>STEPS TOWARD ACHIEVING ATTRIBUTION</u>	25
5.1	Key questions	25
5.1.1	Basic choices for attribution approaches	26
5.1.2	Who decides?	26
5.1.3	Implementing “clean slate” approaches	26
5.2	Social acceptance through incentives	27
5.3	A technical approach to attribution incentives	28
5.4	Reducing the malicious noise floor	30
5.5	Supporting Legal Development	30
5.5.1	Internet attribution is only part of the picture	31
5.6	The need for improved international cooperation	31
5.6.1	Cooperation at a deeply technical level	31
5.6.2	A common multilateral policy framework	32
6	<u>CONCLUSION</u>	34
7	<u>ACRONYMS</u>	37

1 Introduction

The Internet is an information infrastructure enabling global business and government operations, and is simultaneously a central facilitator for free exchange of ideas. While the former function relies on known actors who adhere to rules and behave predictably, the latter function relies on creativity and pure intellect, free of origin. Because the Internet was not designed with either attribution or non-attribution as a requirement, those who need to achieve attribution or non-attribution must constantly evolve to keep up with changing technologies. An average Internet user does not know when attribution or non-attribution is realized, or when to value attribution or non-attribution. Without non-attribution, it is not possible to freely exchange ideas. Without sufficient attribution, it is not possible to enforce policy, law, or treaties to support business and government objectives. Our inability to enforce laws makes creating new ones pointless, and gives malicious actors little incentive to behave.

Cyber attack attribution is a very difficult problem with several unique characteristics:

- Attribution cannot, as we will demonstrate, be accomplished strictly through the use of technology.
- There are situations that make attribution highly desirable, and situations in which attribution would destroy the Internet as a means of open communication.
- Cyber attacks often cross jurisdictional boundaries; hence attribution techniques require cooperation among jurisdictions, some of whom may not be able to trust one another.

Any effective approach to cyber attack attribution must create incentives, making attribution valuable to those who engage in legitimate business transactions and keeping non-attribution possible in the realm of idea exchange. The incentive structure should also penalize those who engage in malicious activities or retribution for idea exchange. The goal of such an incentive structure is to develop social acceptance for the concept of attribution and to guide technological development toward appropriate and sufficient attribution without destroying non-attribution. Over time, incentives should drive the demand for attribution services, resulting in the development of new attribution technologies and approaches. As attribution matures, laws, policies, multi-lateral treaties, and enforcement will evolve. The transition from an Internet free of attribution to one with strong demand for managed attribution services will require significant global investment. Designing attribution services before social acceptance and business demand emerge may be impossible, but we can begin taking steps in the right direction.

Note: Except where otherwise stated, laws and policies discussed in this paper are written from the perspective of the United States (US) but intended for international interpretation, as Internet attribution is a global problem.

1.1 A motivating example

Adequate attribution among mutually-distrusting parties is possible. The nuclear arms race of the Cold War provides a positive example of sufficient attribution. Banning certain nuclear tests and intermediate-range weapons had multi-lateral support for many reasons. While it was advantageous to ban these weapons in a

mutually trusting environment, gaming strategies made it potentially disadvantageous for treaty compliance in a mutually-distrusting environment.

The US government's position was "trust but verify," which was easier said than done. Technology and methods to verify arms reduction or verify and attribute nuclear testing were not available. Through diplomacy, multi-lateral cooperation, technology roadmaps, and a wide variety of processes and procedures, the parties developed means to verify and attribute. Technology alone was inadequate to address the problem. A clear understanding of verification and attribution objectives drove political and technical developments, however, allowing the parties to construct meaningful and enforceable treaties.

Considerable investment is required to achieve enforceable treaties: decades of diplomacy and treaty negotiation, thousands of individuals working together in an international setting to develop technology and procedures, and continuous refinement of treaties and practices. Nuclear non-proliferation treaties can serve as a positive example for managing the reduction of malicious activities on the Internet. Nuclear non-proliferation and managed reduction of malicious Internet activities have several challenges in common: promoting recognition of the problem, achieving international cooperation, developing policies and treaties, developing and enforcing laws, creating enabling technology, and constructing a culture of continuous improvement. At the heart of all of these issues is the need to attribute malicious behavior to an actor or party.

2 What is Attribution?

2.1 Attribution defined

Attribution is defined as determining the identity or location of an attacker or an attacker's intermediary.¹ The focus of this paper is primarily on attribution techniques for attacks via an electronic data network. There are other forms of attack for which attribution is desirable, including physical attack and social engineering attacks. These are important problems, but are outside the scope of this document.

About this paper's definition of attribution:

- Attribution includes the identification of intermediaries, though an intermediary may or may not be a willing participant in an attack.
- Determining motivation, particularly by technical means, is challenging at best; this problem is even more challenging when applied to intermediaries.
- Traceback is frequently used in popular literature as synonymous with attribution techniques. Traceback is "any attribution technique that begins with the defending computer and recursively steps backward in the attack path towards the attacker."² Traceback, thus, is a subset of attribution.

¹ David A. Wheeler and Gregory N. Larsen, *Techniques for Cyber Attack Attribution* (Institute for Defense Analysis, IDA Paper P-3792. October 2003), p. 1

² Wheeler, p. 2.

2.2 Why, and when, attribution is important

The ability to identify the source of a cyber attack is the basis for taking action against the attack's perpetrator. Our legal and policy frameworks for responding to cyber attacks cannot work unless we have adequate attribution; these frameworks remain incomplete because we lack the basis (sufficient attribution) to actually use them. Without the fear of being caught, convicted, and punished, parties ranging from individuals to organizations will continue to use the Internet to conduct malicious activities. We need attribution to create a system of deterrence.

Adequate attribution provides other benefits:

- The prospect of an attacker being identified can serve as a deterrent to future attacks
- Knowing the identity of an attacker, and information gained in the process of attribution, can be used to improve defensive techniques
- Attribution, even partial attribution, can provide the basis for interrupting attacks in progress.

2.3 The difficulty of attribution

A combination of the Internet's architecture and the evolving administrative and governance systems that oversee the Internet make attributing a cyber attack extremely challenging.

2.3.1 Choices of Internet architecture

The telephone system has an effective tracking and billing capability based on the need to charge users of its services on a per-call basis. The Internet's creators never envisioned this need, so the Internet has no standard provisions for tracking or tracing. The Internet model was also not designed to be robust against malicious behavior:

- There are no provisions for cryptographic authentication of information in Internet Protocol (IP) packets. A sophisticated user can modify any information in an IP packet, and, in particular, can forge the source address of a packet (simple for one-way communication).
- Common attack techniques employ a series of stepping stones, using compromised intermediate hosts to launder packets. Transmitted packets can be changed in nature along hops between hosts, so attempting traceback by correlating similar packets is, against a sophisticated attacker, ineffective.

2.3.2 Administrative and policy issues

Unlike the telephone system, Internet attacks can exploit an administrative system that was established when the Internet community was, in essence, a trusted commune, not the global virtual city (with consequent malefactors and little shared sense of community) that it has evolved into.

- Attacks cross multiple administrative, jurisdictional, and national boundaries, with no common framework for cooperation, response, or even trust across jurisdictions.
- The Internet Engineering Task Force (IETF) process does not provide the same global unitary system policy and technical framework that the

International Telecommunications Union (ITU) provides for the telephone system. "There are no universal technical standards or agreements for performing the monitoring and record keeping necessary to track and trace attacks. Moreover there are no universal laws or agreements as to what constitutes a cyber attack, and what punishments, economic sanctions, or liability should ensue. There are no universal international agreements for the monitoring, record keeping, and information sharing necessary to track and trace intruders. No existing privacy laws span the Internet as a whole. Existing international laws and agreements that might touch on these issues were not written for the Internet and need to be tested on cases involving Internet cyber-attacks."³

- High bandwidth means that packet information is not stored for very long. Information useful for tracking is therefore short-lived.

2.3.3 Limited business and economic support

Businesses and government agencies routinely respond to malicious cyber activities by rebooting critical servers, restoring lost data, and identifying and removing the compromise or vulnerability. Today, the fundamental goal of business and government information technology experts is to maintain operations. Because attribution is so difficult, there are very few organizations interested in investigating malicious Internet activity and guiding such investigations through the legal system. Today's protection model, therefore, is primarily aimed at building improved fortifications and attempting to withstand constant attempts to overcome those fortifications.

2.3.4 Other impediments to attribution

Practical concerns that impede effective attribution:

- Tunneling impedes tracking, but is also very useful for creating virtual private networks (VPNs), which are important for security.
- Hackers can often destroy logs and other audit data once they gain system access. This problem is compounded by the fact that system administrators are often poorly trained in security practices.
- Anonymizing services are legitimately valuable to Internet users (e.g., by facilitating political discourse in countries with repressive regimes). While these can be defeated in theory, successful anonymizers demonstrate the practical difficulties of achieving attribution when a sophisticated user desires anonymity.

2.4 What is adequate attribution?

The level to which one seeks attribution can vary significantly, and depends on three interrelated factors: the desired sufficiency of attribution, the nature of the actions for which attribution is desired, and the intended purpose of the attribution.

³ Howard F. Lipson, *Tracking and Tracing Cyber-Attacks: Technical Challenges and Global Policy Issues* (Carnegie Mellon University Software Engineering Institute, Special Report CMU/SEI-2002-SR-009, November 2002), p. 17.

2.4.1 Attribution can take many forms

Sufficient attribution may be satisfied by:

- Knowing the IP address of the host that initiated the attack. This is particularly appropriate if the goal is to modify firewalls to block incoming traffic from these addresses, or so that Intrusion Detection Systems (IDSs) can be programmed to alert on packets from such addresses. One may also be able to discover the machine assigned the IP address at the time of the attributed action. IP address attribution is less useful if the IP address is spoofed or belongs to an intermediate host.
- Identifying the originator's e-mail address. For attacks conducted via e-mail, knowing the e-mail address potentially serves as a useful way to identify the source computer, e-mail address holder, and ultimately the perpetrator. Unfortunately, since e-mail addresses are so easy to create or spoof, an e-mail address is often a dead end for attribution.
- Locating the physical location of the source of the attack, so that jurisdiction can be established, and search warrants or other action be taken, as might be appropriate in the case of crime, terrorism, or war-like activities.
- Identifying the actual individual who was at the attacking computer. As a special variant of identifying the individual, sufficient attribution may also require determining whether the individual was acting on behalf of a foreign government, terrorist organization, or criminal group.⁴

2.4.2 Nature of the attack

One very important distinction is the size of the attack. One of the most devastating forms of attack currently prevalent is the distributed denial of service (DDOS) attack, which has been the focus of much attention, and for which even partially successful attribution can provide important avenues of response.⁵ What distinguishes this, as well as similar forms of attack, is the large volume of attack traffic, which allows probabilistic traceback techniques to play a useful role. Other attacks may involve only a small number of packets; these pose different, and in many ways more challenging problems for achieving adequate attribution.

2.4.3 Intended use or purpose for adequate attribution

Adequate attribution depends on the purpose for which attribution, once achieved, will be used as the basis for action. This involves circular logic: Attribution defines the possible scope of response, while the scope of response is, in turn, an essential part of defining adequate attribution.

Foundational to defining intended purpose of attribution is an understanding of levels of damage or threat (such as what constitutes an act of war in cyberspace).⁶ Unfortunately, there is no established framework for monetizing cyber attack damages or otherwise defining levels of damage or threat, particularly with regard to the international community.

⁴ Rick Aldrich, *Computer Network Defense Attribution: A Legal Perspective* (Prepared for the Defense-wide Information Assurance Program, July 5, 2002), pp.2-3.

⁵ Lipson, 2002,p. 27.

⁶ Lipson, 2002.p. 53.

2.4.3.1 Legal and political needs for attribution

Without reliable attribution, no country can even begin to consider any form of response to a cyber attack. Estonia's experience involving recent denial of service (DOS) and web-defacing attacks against the country's information infrastructure serves as an example. As a North Atlantic Treaty Organization (NATO) member in good standing, Estonia appealed to NATO for help. NATO did not know how to respond and could not determine if cyber attacks are covered by the alliance's treaty.⁷

The lack of attribution for this particular attack has limited Estonia's response to fortification using Internet filtering techniques. Although many have suggested the attacks were in response to the Estonian government's decision to move a statue of a Soviet soldier, the attacks have not been attributed with any certainty. As this contemporary example illustrates, cyber attacks are largely immune to retaliation or punishment, leaving individuals, organizations, and nation-states with little incentive to refrain from conducting cyber attacks out of malice or as retaliation. This example also demonstrates how little incentive exists to modify existing treaties or create new ones when governments have no ability to attribute malicious behavior.

2.4.3.2 As an aid in deterrence through private and legal action

The purpose of law is to proscribe objectionable behavior, make injured parties whole, and end disputes.⁸ However, most legal systems lack the capacity to respond to every infraction and most societies lack tolerance for repeated violations. Therefore, a basic function of law is to create incentives that deter malicious behavior. In order for such a deterrence to work, the perceived risk of engaging in malicious behavior must exceed the perceived benefit. According to Jim Gosler, a Sandia National Laboratory Fellow, cyber actors engaged in malicious activity perceive risk through a combination of factors: the probability of detection, the probability of attribution, the penalty for getting caught, and the cost to act.⁹ Note that this model is a work in progress, but, as a general rule, risk tends to increase as any of the factors increases. If the probability of detecting malicious activity is zero, the other factors are irrelevant in formulating a risk value. If an objectionable activity is detected and the probability of attributing the malicious behavior is zero, the offending party is not at risk. Under these circumstances, penalties have no relevance to an offending party and there exists no disincentive to malicious behavior. Laws are only enforceable if the justice system can attribute a crime to an actor with sufficient confidence to respond. Today, Internet users can break established laws through the Internet with little or no fear of retribution because current attribution techniques are insufficient. Internet fraud increased by 132% from 2004 to 2005.¹⁰

Adequate attribution, once achieved, can serve as the basis for private or legal action:

- Most insider threats, once detected, are handled internally by the organization affected, though a variety of laws (particularly the Computer Fraud and Abuse

⁷ Anne Applebaum, *For Estonia and NATO, A New Kind of War*, Washington Post, Tuesday, May 22, 2007

⁸ *Business Law*, Emerson, 1997, p. 4

⁹ James R. Gosler, *Transforming US Intelligence*, ed. Jennifer E. Sims and Burton L. Gerber (Washington, DC: Georgetown University Press, 2005), pp.106-107.

¹⁰ Internet National Fraud Information Center, *2005 Internet Fraud Report*, http://www.fraud.org/2005_Internet_Fraud_Report.pdf (December 2007).

- Act, USC 18 Section 1030) are directly applicable.¹¹ Insider threats, once detected, are generally removed, blacklisted, or punished some other way.
- US law has a well established framework for establishing adequate attribution. Some key concepts include:
 - Proof is “the effect of evidence, the establishment of fact by evidence.”¹²
 - Evidence is “a narrower term, and includes only such kinds of proof as may be legally presented at a trial, by the act of the parties, and through the aid of such concrete facts as witnesses, records, or other documents.”
 - Proof beyond a reasonable doubt is “such proof as precludes every reasonable hypothesis except that which it tends to support and which is wholly consistent with defendant’s guilt and inconsistent with any other rational conclusion.” This is the legal standard required in criminal cases.
 - Probable cause is “reasonable cause; having more evidence for than against ... A set of probabilities grounded in the factual and practical considerations which govern the decisions of reasonable and prudent persons and is more than mere suspicion but less than the quantum of evidence required for conviction.”

The standard of proof is higher for criminal action than for civil action. Hence, within the legal system, there is a distinction, depending upon intended action, as to what constitutes sufficient attribution.

2.4.3.3 As a basis for war and response to cyber terrorism

Establishing the basis for war or armed response as a result of cyber attack from either a nation-state or terrorist group is perhaps the least well defined aspect of the role of attribution as a basis for response. This question can be viewed from three perspectives: what is generally considered the law of war, Constitutional and legal authorities of the US, and established international conventions, particularly those of the United Nations (UN) and NATO. The key issue is that the option of resorting to war in the face of cyber attack remains a contentious and undefined topic. How and in what form a cyber terrorist attack would properly merit response is an even less well defined issue.

Traditional just war theory operates from a “will to peace” and is traditionally divided into ethical considerations of just recourse to war (*jus ad bellum*) and just conduct in war (*jus in bello*). According to the Christian or Western just war tradition, *jus in bello* actions are those that are discriminate and proportionate. “Discriminate” means that destructive actions are aimed at the adversary and that non-combatants and the innocent are not targeted. “Proportionate” means that the effects of our destructive actions are not out of proportion with the ends we seek to achieve. Justice, therefore, does not exist in a vacuum: Justice, discrimination, and proportionality

¹¹ Jeffrey Hunker, Taking Stock and Looking Forward, in *Insider Attack and Cyber Security – Beyond the Hacker* (Springer, 2008).

¹² *Black’s Law Dictionary*, 6th edition, also same reference for subsequent definitions.

only have meaning with respect to that which threatens peace and makes decisions about war necessary.¹³

The UN Charter is the acknowledged mechanism for determining the lawfulness of the resort to force by nations. Key provisions are:

- Article 51, which recognizes each state's inherent right to self-defense against armed attack
- Article 2(4), prohibiting the threat or use of force against the territorial integrity or political independence of any state
- Article 39, which permits the Security Council to identify and label an event as a "threat to the peace, breach of peace, or act of aggression"¹⁴
- Article 42, which states that "should measures [such as economic sanctions, or diplomatic actions] be inadequate or have proved to be inadequate, it may take such action by air, sea or land forces as may be necessary to maintain or restore international peace and security."¹⁵

Hence, there are two exceptions to 2(4): a use of force pursuant to a mandate issued by the Security Council in accordance with Article 42, and acts of self-defense consistent with Article 51.

There are at least three different interpretations of these provisions relevant to response to an attributed cyber attack. The first, common among military operators and decision makers, is that the "use of force" prohibition seeks to keep incidents that are below a certain threshold of violence from mushrooming into full-blown wars; it is not the method of attack that matters, but the amount and type of damage done. The second takes the position that the Charter was meant to favor resolution of conflict by non-military means. Consistent with this approach, only an armed attack (a classic attack with traditional military forces) constitutes a use of force. Therefore, according to this definition, the method of attack is what matters. A third, even less well defined interpretation, is that response merits a case-by-case analysis.¹⁶

Whether and how this collective body of law applies to information warfare remains an unsettled question, since most of this law far predates computers. Most legal commentators would agree that it does apply to information warfare, though as recent examples (e.g., Estonia) suggest, its application has yet to be tested.¹⁷

2.5 When attribution is not desirable

The importance of non-attribution in protecting ideas and minority views from oppressive organizations or regimes cannot be overstated. The Internet has become a very effective medium for freely sharing ideas. For many, the fear of retribution makes it one of the only methods for freely expressing ideas. Mechanisms developed to facilitate attribution must enforce non-attribution for idea sharing. Well-crafted attribution mechanisms need to make it easy to attribute malicious behavior (as

¹³ T.K. Kelly, , *The Just Conduct of War against Radical Islamic Terror and Insurgencies*. Monograph. 2007.

¹⁴ Thomas C. Wingfield, ,*When is a Cyber Attack an "Armed Attack?":Legal Thresholds for Distinguishing Military Activities in Cyberspace*(Cyber Conflict Studies Association, February 1, 2006), pp. 4-6.

¹⁵ UN Charter

¹⁶ Wingfield, pp. 6-7.

¹⁷ Aldrich, p. 6

defined by society through a legal/policy process), nearly impossible to attribute freely exchanged ideas, and possible to attribute retribution.

2.6 A purely technical solution is impossible

To better understand why technology alone cannot provide sufficient attribution, consider an ideal network with perfect attribution. Such a network does not exist, but can help in illustrating the limitations of technology in accomplishing our attribution goals. The perfect attribution network (PAN) is designed and constructed so that all actions taken by a specific user on a particular machine are fully attributable to that machine and user.

The foundation of the PAN provides the attribution services and cannot be altered or bypassed by any user or administrator. Any application installed on the PAN interfaces with the attribution foundation and adopts a complete set of attribution services. It is impossible for an application to bypass the attribution services or to alter the services in any way. Moreover, applications, such as online banking, installed on the PAN do not require modification to invoke the attribution services. The purpose of developing the PAN model in this way is to show that even perfect technical attribution services can be defeated.

Now consider an application installed on the PAN by a community of users wishing to engage in non-attributable actions within and outside of that community. Each instance of the non-attribution application (NAA) can communicate with every other instance of the NAA. While every point-to-point message processed by the NAA is fully attributable to the source and destination users by the underlying PAN, the attribution scope is limited to the immediate source and destination of each message.

One strategy to achieve non-attribution is to remove point-to-point context from all messages through the application of an NAA overlay (NAAO). As a simple example, consider an NAAO configured in a logical ring topology. Messages in this ring topology flow clockwise and each NAA instance receives all of its incoming messages from a single NAA instance (the instance immediately counterclockwise) and transmits all outgoing messages to a different single NAA instance (the instance immediately clockwise). The NAAO provides strong confidentiality of all messages, generates random messages to maintain a roughly constant bandwidth of data flow irrespective of actual message content, and abstracts the actual message source and destination from the PAN. Direct PAN features cannot identify which messages are authentic, the actual source or destination of any given message, or the actual source of a message leaving the NAAO. Further, constant bandwidth utilization in the ring topology makes traffic analysis very difficult (not a specified function of the PAN).

This simple example demonstrates that a purely technical attribution solution is not possible. It is possible to develop features that facilitate the attribution process, such as unique communication keys, traceback, and logging. However, these solutions cannot be relied upon to provide sufficient attribution for most cases of malicious behavior. Depending on the severity of malicious conduct, options apart from the information infrastructure will be required.

3 Attribution Policy

3.1 Policy challenges to attribution

There are some unique challenges to successful attribution that shape the possible range of steps toward more effective attribution.

3.1.1 Incentives for Internet-wide solutions

Installing new capabilities on routers is a special case of a larger challenge. Put bluntly, the record of adopting Internet-wide of new protocols has not been good. This problem is largely due to the new capabilities that some proposed solutions would require from routers, but the problem is of a more general nature. Some specific examples include:

- IPv6
- Domain Name System (DNS) Security Extensions (DNSSec)
- Modifications to the Border Gateway Protocol (BGP).

These cases all have two common underlying issues: Effectiveness of the proposed changes requires uniform adoption of the new protocol, and the costs and burden accrue to the individual entity, while the benefits are distributed and system-wide. Unlike the telephone system, for example, in which the ITU has an effective mechanism for creating and enforcing technical requirements, the request for comment (RFC) process of the IETF is essentially voluntary. Hence, a major policy issue is how to create an incentive or regulatory system that, as appropriate, would ensure that system-wide changes are implemented.

3.1.2 Establishing trust across administrative or jurisdictional boundaries

Within a unitary administrative jurisdiction, a level of trust can be reasonably assumed. Across administrative domains, however, trust—along with the requisite levels of cooperation and collaboration—cannot be assumed. Trust ultimately depends on particular relationships among individuals, and their roles within organizations and groups. Trust cannot be established among anonymous actors.

3.1.3 Balancing privacy and attribution

Privacy—or the expectation of anonymity—is an important social expectation for many Internet users. Perhaps even more important are the political and social freedoms that are protected by right in countries like the United States, but suppressed in oppressive regimes. Internet anonymizing capabilities are an important way of advancing personal and political freedoms and cannot be ignored.

“Trust and privacy trade-offs are a normal part of human social, political, and economic interactions, and such trade-offs can be resolved in a number of venues, particularly in the marketplace.”¹⁸ In “Caller ID”, for example, a marketplace has emerged for balancing privacy and attribution. A customer can pay for attribution services in the form of Caller ID. A customer can also pay to have additional privacy, blocking his/her identity on the phones of customers who use Caller ID. The Caller ID customer, in turn, can choose to block incoming calls from anonymous callers; the blocked caller then has the choice of removing the block feature for that call. In this system, there is a form of negotiation in which mandates play no part. The caller can

¹⁸ Lipson, p 58

choose to relinquish a degree of privacy in order to complete the call; the recipient can choose whether to accept anonymous calls.¹⁹

3.1.4 Liability

Liability for the causes and consequences of cyber attacks is an undeveloped field within US jurisprudence, and all of the questions that apply more generally to liability for domestic-only cyber attack can with equal applicability be applied to issues of liability internationally. There really are no clear answers to cyber attack liability questions, and relevant case law is currently lacking.

Among the key questions are:²⁰

- What are the liability exposures for:
 - Perpetrators of the attack?
 - Vendors of software that made the attack possible?
 - Owners and administrators of the intermediate (e.g., zombie or anonymizer) systems that participated in generating attack packets or obscuring the original source of the attack?
 - Transmitting Internet Service Providers (ISPs) and networks that did not squelch the attack when notified or did not help trace the attack in accordance with international policy agreements (to the extent they exist)?
- Would there be waivers of certain kinds of liability exposure for those who participate in tracking and tracing?
- Should liability extend to the owner of a system that participated in attacks without the owner's knowledge or permission?
- Are those who provide anonymizer services liable for providing anonymity to attackers, even inadvertently?

Liability will eventually form an important component in the policy framework for cyber security generally, and that framework will affect the range of feasible options for effective attribution.

3.1.5 Evidence preservation

Related to the security of attribution mechanisms, but a distinct issue, is that in forensic analysis, determining attribution may require tampering with the evidence. Preservation of the chain of evidence, and the extent to which evidence can be tampered with, will depend both on the context and purposes of the attribution effort.

3.1.6 Administrative context of attribution activities

There are several important distinctions regarding the administrative context of attribution activity. These can be thought of as distinctions based on the extent to which attribution efforts cross jurisdictional and administrative lines, the specific organizations involved in the attribution efforts, the timing of the needed attribution, and the extent of repositioning. Many of these distinctions raise complex issues:

¹⁹ Ibid

²⁰ These questions are drawn from Lipson, pp.53-54.

- Though perhaps self-evident, once attribution requires crossing jurisdictional lines (whether legal or organizational) new attribution challenges emerge.
- An important distinction is whether the attribution is taking place in the context of national security, or in commercial or personal realms. The distinctions among which organizations are involved are important: Each type of organization faces different legal obligations and proscriptions, may play multiple roles, and has different incentives for action and cooperation.
- It may be useful to achieve attribution long after an attack takes place, but in many cases attribution must be achieved as soon as possible to be of value. Understanding the required timeframe for attribution is important in terms of the technologies and systems necessary to achieve attribution, as well as in determining the adequacy of various legal and policy instruments.
- The options available for adequate attribution will vary depending upon the extent to which attribution mechanisms have been prepositioned. Such prepositioning has a number of dimensions: technical capabilities, administrative arrangements and understandings, legal frameworks, and international agreements.

3.2 Domestic legal authorities

A number of different laws both allow and proscribe actions aimed at attribution. This area is particularly dynamic: Both the letter and interpretation of the USA PATRIOT Act, for example, are, at the time of this writing, in a state of flux.

3.2.1 Interception of communications

The Federal Wiretap Act (18 USC 2511) generally prohibits interception of telecommunications or computer communications without a specific court order, either a Title III court order (for law enforcement) or a Foreign Intelligence Surveillance Act (FISA) court order. The distinction among court authorities is important: Sometimes it may be critical to determine whether the perpetrator was acting on behalf of a foreign power or a terrorist organization so that a FISA warrant may be sought.²¹

There are important exceptions to the Federal Wiretap Acts prohibitions:

- Service providers may be able to rely on the consent exception by requiring users to sign user agreements or click through consent banners. Consent banners may establish implied consent for monitoring communications.²²
- There is a service provider exception to the general prohibition against interceptions provided under the Federal Wiretap Act, so that for a service provider it is not illegal “to intercept, disclose, or use that communication in the normal course of his employment while engaged in any activity which is a necessary incident to the rendition of his service or to the protection of the rights or property of the provider of that service.”²³

²¹ DIAP, , p. 7

²² Wheeler, p. 4

²³ 18 USC 2511 (2)(a)(i)

3.2.2 Cooperation among law enforcement and intelligence services

Until the passage of the USA PATRIOT Act, law enforcement agents could not even try to identify a hacker who had illegally penetrated a government-interest computer (which includes most domestic computer systems) without first obtaining the hacker's consent or a court order. The USA PATRIOT Act made two important changes:

- It established a new exception for intercepting the communications of "computer trespassers"
- It permitted the increased sharing of information between law enforcement and intelligence agencies.²⁴

3.2.3 Limitations on intelligence activities

Activities by intelligence agencies are limited by the Fourth Amendment (at least regarding activities within the United States or against United States persons, as defined in Executive Order 12333)²⁵.

3.2.4 Cooperation of intermediate ISPs and systems administrators

A number of attribution techniques require the cooperation of intermediate systems administrators (e.g., traceback). If the perpetrator is not one of the system's own subscribers, system administrators are not prohibited by law from voluntarily cooperating with law enforcement. If the suspected perpetrator is a subscriber, and the electronic communications service provider does not provide service to the public, the ISP could voluntarily choose to provide information to the investigating authorities. If the ISP provides services to the public, or a non-public ISP chooses not to voluntarily cooperate, then the investigating authority has two options:

- Provide a federal or state search warrant (based on 18 USC 2703 (d)). However, obtaining a warrant or court order may take some time, and ISPs are not required to keep historical transactional data for any specific period of time (or at all).
- Request the ISP to preserve the information for 90 days (with an additional 90 day extension possible) under 18 USC 2703(f).

Unfortunately, while 2703(d) orders can be issued with directions prohibiting the ISP from divulging information related to the government's request to unauthorized parties, no such protection exists under 2703(f); hence if there is a concern that the ISP is complicit in the attack, it may be necessary to expedite a warrant without the 2703(f) request.²⁶

Once it has been determined that the perpetrator is a subscriber to a particular ISP, obtaining subscriber information generally requires an administrative subpoena, a federal or state grand jury or trial subpoena, or, in the case of an intelligence investigation, a National Security Letter. There are limited exceptions: When the subscriber consents, or when the service provider reasonably believes that an emergency involving immediate danger of death or serious bodily injury to a person justifies disclosure without delay.²⁷

²⁴ DIAP, pp. 7-8.

²⁵ Ibid, p. 8.

²⁶ Aldrich, p. 7.

²⁷ Ibid, p. 8.

3.3 International cooperation

In many cases, attribution will require the cooperation of international entities; in those cases, there are a number of avenues for seeking attribution:

- Informal law enforcement requests for providing information are probably the most commonly used means for obtaining information.
- The Council of Europe Convention on Cybercrime was signed by 33 countries in December 2001; the US ratified the agreement in 2007. It provides a general framework for harmonizing laws against computer trespass, and for expedited cooperation among the signatories for investigation of computer crimes. It is, however, a general framework agreement, and provides for no specific mechanisms or approaches for attribution.
- Mutual Legal Assistance Treaties exist between the US and roughly a dozen other countries. These allow the US to serve a subpoena in a foreign country by making a request through the US State Department. A 6-month delay by a foreign country in honoring the request is not unusual. Moreover, each request expends a certain amount of political capital, so requests are reserved for major cases only.²⁸
- Letters rogatory are a means by which a court in one country formally requests the assistance of a court in another country to examine witnesses or otherwise assist in the administration of justice.
- In a well-publicized sting operation, the Federal Bureau of Investigation (FBI) lured two Russian hackers to the US by offering them consulting jobs that appeared to be genuine; the FBI then used information given to them by the hackers (who were demonstrating their skills) to download information from the Russian computers for use as evidence. Both men were subsequently arrested and indicted. Ironically, Russia's Federal Security Service later charged one of the FBI agents with illegally accessing the hackers' computers, which were physically located in Russia. The FBI called this the first case in its history to "utilize the technique of extra-territorial seizure." From the FBI's perspective, immediate access to the evidence on the Russian computers was essential for pursuing the case; from the Russian perspective, traditional national boundaries and legal jurisdiction were violated, setting a troubling precedent.²⁹

There are models of effective international cooperation at a technical level that might provide useful models for cyber attack attribution—specifically, the Financial Action Task Force (FATF), established by the G-7 in 1989 to help deal with the problem of money laundering. Money laundering exploits the mismatch between laws based on national boundaries and international money flows that cross those boundaries. The FATF has assessed the effectiveness of existing cooperative efforts to "prevent the utilization of the banking system and financial institutions for the purpose of money

²⁸ Lipson, p. 51.

²⁹ Ibid.

laundering” and has conducted many technical studies to improve awareness and multilateral cooperation of member nations.³⁰

4 Attribution Technology

4.1 Technical challenges to attribution

4.1.1 Entry point anonymity

Even if an attack packet can be attributed to the IP address of its host of origin, it is becoming increasingly difficult to link that IP address with the actual perpetrator of the attack. There are a number of ways in which the physical identity of the perpetrator is now decoupled from the IP address:

- Prepaid Internet address cards, in which there is no requirement for personal identification by the ISP
- Cyber-cafes
- Public Internet access in libraries and other facilities.³¹

4.1.2 Retention capabilities of routers

Two separate issues come together to create a challenge to attribution:

- Storage intervals on routers, particularly those routers in or near the high-speed core of the Internet, are very short. Hence, forensic techniques either have to be very rapid (before the router cache is overwritten) or new capabilities have to be created aimed at preserving routing information.
- Increasingly sophisticated attacks mean that successful attribution will have to deal with attacks that are extremely rapid as well as with attacks that are extremely slow (possibly over a period of months).

4.1.3 Securing attribution mechanisms

Attribution techniques need to be secured against attack or subversion. This presents several challenges:

- Attribution techniques and technologies need to be protected, particularly software used in authentication efforts
- Data used for attribution needs to be protected
- Attribution techniques must not create new avenues of exploitation (e.g., by creating a new technique for performing a DOS attack against the system³²)
- Achieving adequate security of attribution mechanisms often requires establishing and maintaining trust across jurisdictions.

4.2 Selected current technical approaches to attribution

There are a number of alternative approaches, both in use and proposed, for providing attribution to cyber attack. The table below summarizes some of these.³³

³⁰ Wolfgang H. Reinicke, *Global Public Policy: Governing without Government?* (Washington, DC: Brookings Institution Press, 1998), pp. 157-172, as cited in Lipson, p. 48.

³¹ Lipson, p. 56.

³² Wheeler, op. cit. p 48.

³³ This table is based on Wheeler, pp. 10-11.

Technique	Description
Hash based IP traceback	Routers store a hash (relatively unique, compressed representation created by a one-way function) of each packet across the network; attribution is done by tracing back the hash across network routers.
Ingress filtering	Require that all messages entering a network have a source address in a valid range for that network entry point. This limits the range of possible attack sources.
ICMP return to sender	Reject all packets destined for the victim; return rejected packers to their senders.
Overlay network for IP traceback	An overlay network links all ISP edge routers to a central tracking router; hop-by-hop approaches are used to find the source.
Generating trace packets using control messages (e.g., iTrace)	Periodically (e.g., 1 in 20,000 packets) a router sends an ICMP traceback message to the same destination address as the sample packet. The destination (or designated monitor) collects and correlates the tracking information.
Probabilistic packet marking	A router randomly determines whether it should embed information about the message's route into a given message. The defender can then use a set of messages to determine the route.
Hackback	Insert querying functionality into a host, specifically without the permission of the owner. If an attacker controls the host, this may alert the attacker and make the information less reliable.
Honeypots	Decoy systems that are only accessed by attackers capture information for attribution.
Watermarking	A passive technique that brands a file as belonging to a rightful owner.

4.2.1 Hash-based IP traceback

In theory, tracking traffic can be addressed by keeping a log at each router in the Internet of every packet that passes through it in a tamper-proof, fully-authenticated manner. While intuitively appealing, this approach is impossible in practice; it would require virtually unlimited fast storage at each router. In addition, this would raise both privacy and security issues:

- Privacy would be impossible to maintain in an environment in which complete traffic logs, including all packet contents, were kept at all routers in the Internet
- Implementing the requisite router capabilities Internet-wide would present implementation incentive issues already noted
- Security of the logs themselves would be an issue; obviously these logs would present tempting targets to attackers
- Trust across jurisdictions would have to be established.

Some of these issues have been addressed through hash-based IP traceback, in which a hash of each packet would be stored instead of the full packet itself. This greatly reduces the storage requirements, and addresses the privacy issues, since only the packet digests are stored at each router, and not the actual packet contents. Legally, acquiring access to hashes appears akin to "trap and trace" techniques applied to telephone communications, although it is unclear if the courts will agree with this viewpoint.³⁴

During its transit through the Internet, a packet can be transformed in a number of ways, including address translation, tunneling, and fragmentation. Transformation information corresponding to packet digests can be stored in a transformation lookup table. Consequently, this transformation information, correlated with the hash, would be a necessary part of any attribution system using hash-based traceback.³⁵

³⁴ Wheeler, p. 13, and Lipson, pp. 44-46

³⁵ Lipson, pp. 44-45.

4.2.2 Network ingress filtering

This approach restricts network traffic by requiring that each message entering a network have a valid source address for that network entry point. This approach could be applied to protocols other than IP. The occurrence of packets with spoofed source addresses and their ability to transit the Internet can be greatly limited through cooperative efforts by ISPs using network ingress filtering. To limit IP source address spoofing, the ISP places an ingress filter on the input link of the router that carries packets from the customer network into the ISP's network and onto the Internet. The ingress filter is set to forward along all packets with source addresses that belong to the known set of IP addresses assigned to the customer network by the ISP, but the filter discards (and optionally logs as suspicious) all packets that have source IP addresses that do not match the valid range of the customer's known IP addresses. Hence, packets with source addresses that could not have legitimately come from within the customer network will be dropped at the entry point to the ISP's network.³⁶

The widespread use of ingress filtering by all service providers would greatly limit the ability of an attacker to generate attack packets using spoofed source addresses, making tracking and tracing the attacker a much easier task. Network ingress filtering has a number of advantages:

- It is easily implemented using existing infrastructure
- It supports attribution without requiring message logs of unrelated messages or additional network bandwidth
- It has no known legal impediments.³⁷

Major disadvantages:

- There is a broad element of corporate citizenship involved; many of the benefits of ingress filtering would accrue to potential attack victims who are not the filtering ISP's customers.
- Network ingress filtering is primarily useful for internal network attribution, and to determine if an attack came from the outside.
- Ingress filtering can break certain network services, and ISPs would have to expend extra effort to get those services to work despite ingress filtering.³⁸

In the longer term, this approach could be applied nationwide or internationally. The benefit would be to limit potential sources of attack, though attacks originating in non-participating jurisdictions would be immune to the attribution provided by this technique.

4.2.3 ICMP return to sender

With the prevalence of DDOS attacks, a number of techniques have been developed to trace large numbers of packets back to their entry points in an administrative network domain. With this technique, which is applicable only while a DOS attack is underway, the ISP first configures its routers to reject all packets destined for the victim. Rejected packets are returned to their senders: An Internet Control Message Protocol (ICMP) "destination unreachable" error message packet is sent back to the

³⁶ Lipson, p. 30

³⁷ Wheeler, pp. 34-35.

³⁸ Wheeler, pp. 35-36.

source IP address listed in the rejected packet. Analysis of these ICMP messages allows for rapid identification of the routers serving as the entry points at the outermost boundary of an ISP's network. The ISP can then identify the specific router interfaces (the peering point, or connection to a neighboring ISP) through which the attack is entering its domain; the neighboring ISPs can then continue tracing the attack back, using this or other techniques.

This technique is a fast and efficient way to trace current DDOS attacks to the boundary of an administrative domain. Its drawbacks as an attribution technique are:

- It depends heavily on the specific characteristics of the DDOS attacks it was intended to defeat (it requires attack packets have source IP addresses randomly distributed throughout the Internet address space, including IP addresses that are invalid because they have not yet been allocated)
- It depends on a large number of attack packets being directed at the target
- For ultimate attribution, it depends on the trustworthiness, cooperation, and skills of upstream ISPs that lie between the attack target and the attack source.³⁹

4.2.4 Overlay network for IP traceback⁴⁰

This approach (particularly as proposed in CenterTrack⁴¹) improves the traceability of large packet flows associated with DOS attacks by creating an overlay network, using IP tunnels to connect the edge routers on an ISP's network to special-purpose tracking routers that are optimized for analysis and tacking. The overlay network is designed to simplify hop-by-hop tracing by having only a small number of hops between edge routers. In the event of a DOS attack, the ISP diverts the flow of attack packets destined for a victim's machine from the existing ISP network into the overlay tracking network; the attack packets can then be traced back, hop-by-hop, through the overlay network, from the edge router closest to the victim back to the entry point into the ISP's network.

This approach has not been implemented (and is referred to as a defunct research project⁴²), as it has a number of flaws:

- It results in an increase in overall network complexity, which can lead to operational errors (e.g., in routing updates)
- The overhead inherent in creating IP tunnels could amplify the negative effects of a DOS attack
- The technique does not work with attacks that originate inside an ISP's network
- The approach may not be scalable for distributed DOS attacks with many entry points into the target ISP's network
- It requires cooperation across jurisdictional domains.

³⁹ Lipson, p. 37.

⁴⁰ This description is based on Lipson, pp. 39-40.

⁴¹ Robert Stone, *CenterTrack: AN IP Overlay Network for Tracking DoS Floods* (9th USENIX Security Symposium, Denver, Colorado, August 14-17, 2000), pp. 199-212.. Available at http://www.usenix.org/publications/library/proceedings/sec2000/full_papers/stone/stone.pdf

⁴² A comment on a website. See <http://staff.washington.edu/dittrich/misc/ddos>

4.2.5 Probabilistic packet marking

A router randomly determines whether it should set information about the message's route into a given message; the defender can then use a set of messages to determine the route. This approach involves placing tracking information directly into rarely-used header fields inside IP packets. In most cases a number of messages must be received before attribution can be achieved. This approach does not require interactive operational support from other ISPs, and can be employed after the fact.

A disadvantage is the potential for an attacker tampering with, or spoofing, the tracking information, but the tracking information can be subject to authentication.⁴³

4.2.6 Generating trace packets (iTrace)⁴⁴

For this technique, routers send separate messages that can be used to support attribution; under the iTrace system, traceback messages are sent occasionally (e.g., 1 in 20,000 packets).⁴⁵ The destination (or monitor) collects and correlates the tracking information. For large packet flows, sufficient information can be collected to successfully trace the attack.⁴⁶

This approach poses a number of problems:

- Since messages supporting tracing may be routed separately from the message being traced, extra effort is required by any implementation to associate the trace message with the message being traced.
- An attacker can defeat or disrupt the trace by sending spoofed iTrace packets, iTrace packets should include an authentication field. Which authentication method should be used is an open research question, based on a trade-off between cryptographic strength and computational resources. If attackers make their attacks look like trace messages, network resources could be overwhelmed.
- There is a public key infrastructure issue as to who has the right to sign an iTrace packet, and how one validates that signature.

4.2.7 Hack-back⁴⁷

In its benign form, services that assist in providing needed attribution information can be added to the system after an attack has been detected. In its most extreme form, hack-back involves breaking into a host machine or a series of host machines, attempting to go backward toward the attacker. Reportedly the Air Force has used this approach, calling it "Caller ID", to track down and arrest an intruder.⁴⁸ "Caller ID" is based on the belief that if an attacker goes through intermediate systems in the process of an attack, there is a high probability that the intermediate systems have known vulnerabilities. The defender, knowing the same attack methods as does the attacker, can simply reverse the attack chain.

⁴³ Song, Dawn, and Perrig, Adrian. "Advanced and Authenticated Marking Schemes for IP Traceback." 878-886. IEEE Infocom 2001. See <http://www.ieee-infocom.org/2001/>

⁴⁴ This description based on Lipson, p. 41 and Wheeler, p. 20-21.

⁴⁵ See the IETC ICMP Traceback Working Group website: <http://www.ietf.org/html/charters/itrace-charter.html>

⁴⁶ Lipson, p. 41

⁴⁷ This description based on Wheeler, pp. 23-24 and Lipson, pp. 29-31.

⁴⁸ Wheeler, p. 23

There are a number of disadvantages to this approach:

- Any attempt to insert a host monitoring function may be noticed and/or countered by the attacker.
- Especially if the hack-back is performed by anyone other than the host owner or authorized administrator, there are serious privacy violations that may be incurred.
- Because this process involves unauthorized entry into intermediate ISPs to obtain routing information, it would likely be deemed illegal as a violation of the Computer Fraud and Abuse Act (18 USC 1030); however, section (f) provides a potential exception for law enforcement and counter-intelligence agents, allowing for "lawfully authorized investigative, protective or intelligence activity."⁴⁹ The Computer Crime and Intellectual Property Section of the Department of Justice has taken the position that hack-backs cannot be lawfully authorized for law enforcement or intelligence agents.

4.2.8 Honeypots

Honeypots are systems that appear normal but, in fact, are never accessed by normal users. For attribution, honeypots can reveal attack paths in ways that an attacker may not anticipate.⁵⁰

Both operational and legal issues are raised by honeypots:

- Honeypots only work if the attacker chooses an attack path through the honeypot.
- Honeypots require monitoring and analysis, which has both resource and expertise demands.
- Domestic and international agreements need to clarify whether breaking into a honeypot would be sufficient cause for retaliation (legal or otherwise), or would instead be considered entrapment. From the perspective of a possible multilateral system for attribution, there would be need to determine whether the system should be triggered by attacks against honeypots, or should be reserved for real attacks.
- As Aldrich notes, "this technique raises a myriad of legal issues for system providers, law enforcement, and counterintelligence personnel. The warfighter may be less inclined to use this technique because of its potentially slow and reactive methodology. To the extent such use was envisioned, it seems likely that compliance with the requirements for intelligence personnel would be observed absent exigent circumstances for which reliance on the President's Commander in Chief role would be deemed the only timely option."⁵¹

⁴⁹ DIAP, p.12.

⁵⁰ Wheeler, pp.27-28.

⁵¹ DIAP, p. 12

4.2.9 Watermarking

Unlike honeypots, in which executable code can be used to facilitate the transmission of an attacker's identity, watermarking is a passive technique that relies on a concealed unique marker identifying a file as belonging to a rightful owner. This enables later confirmation that a suspect is truly the attacker because the data is made unique through watermarking. However, this requires prepositioning of watermarks on selected files.

The legal issues primarily concern where attribution efforts can be directed to search for watermarked files. The law is reasonably clear that data one exposes to the public has no reasonable expectation of privacy; hence files posted publicly (e.g., on Web sites) could be searched. There is an argument that even files not open to the public should still be legally open to using technologies that search out watermarks. The software agent used to track the watermarked file would see only that which the attacker was not entitled to have in the first place; therefore it would be minimally intrusive and maximally effective. However, "even as compelling as this logic seems, it would require a rethinking of how the Fourth Amendment has been applied traditionally."⁵²

A variant of watermarking is called "sleepy watermark tracing": The defender seeking attribution injects a watermark into the return data flow. This approach depends on the attack protocol being bi-directional, with data flowing back to the attacker (or intermediary). One advantage of the technique is that it can attribute immediately through a large number of stepping stones, provided the data is not transformed (e.g., encrypted). However, sleepy watermark tracing would require significant changes to existing systems. Data detectors for the reverse flow must be placed in locations that can actually observe the data, and detectors may report false positives. Also, hosts that transform the data may foil the technique.⁵³

5 Steps Toward Achieving Attribution

This section attempts to answer a very hard question: "Given that we need attribution, but that in general attribution is impossible, what should we do?" Our approach is to suggest steps in the right direction rather than outline a plan. We begin by raising some key questions, such as, "Who should decide when attribution is warranted?" Next, we describe the need for incentives to modify Internet users' service expectations and gain an understanding of how attribution can provide positive value. Finally, we suggest a progression of steps: lower the noise floor of malicious activity, support Internet legal development, and establish the basis for multilateral treaties and international response to malicious activity.

5.1 Key questions

Having reviewed existing approaches, either implemented or proposed, it is important to step back in order to recognize that there are some important questions that need to be addressed, but that tend to be obscured when viewed in relation to proposed solutions.

⁵² DIAP, p. 12

⁵³ Wheeler, p. 42

5.1.1 Basic choices for attribution approaches

In considering approaches to attribution, a great deal of attention is focused on specific techniques, but hardly any at all on what sort of system we want to have. For example, we could:

- Have a system like the telephone system, in which attribution is automatic (back to the originating number); a user can turn the feature off, but that decision can be overridden by law enforcement/courts.
- Have a system in which users can opt-in or opt-out without any subsequent recovery ability by the system.
- Have multiple networks, some provide varying levels of attribution, others none at all. Users decide what networks to accept packets from.

5.1.2 Who decides?

Attribution needs to be thought of as a system:

- There is the process of providing attribution
- There is the decision as to what constitutes sufficient attribution
- There is the response to attributed behavior.

Which cyber activities need attribution? Which cyber activities need non-attribution?

There may be an issue of Internet governance here: Just as a court decides when a wiretap is appropriate, someone must decide when to use attribution, assuming that it is not automatically employed, and someone must decide when to act upon the attribution. As noted earlier, there is an extensive canon for determining appropriate responses (e.g., to take military action, impose sanctions, or seek legal redress). But attribution may require the cooperation of multiple entities, and that cooperation may be conditioned upon having some say in the eventual outcome. Further, whatever the existing canon, it remains untested with regard to response to cyber attack.

5.1.3 Implementing “clean slate” approaches

The Internet was not designed for the purposes it is now used for, nor for the security demands created by a user base without shared trust. A number of “clean slate” network projects are now underway, including the National Science Foundation (NSF) Future Internet Network Design (FIND) project and the Department of Defense (DoD) Global Information Grid (GIG). These projects’ goal is to develop a new network of networks that both addresses the current needs for which the Internet was not originally designed and anticipates capabilities for future needs.

The challenge to such projects is that network architectures are not just about technology; architecture includes the social, legal, administrative, and economic infrastructures in which the network architecture operates. Secure BGP and IPV6 have faced slow acceptance not just for technical reasons, but also because these other factors play a role.

We do not now have a plan for adoption of a “clean slate” network; if some of the current “clean slate” work is successful, we may be, in colloquial terms, like the dog that, after chasing a car, finally catches it. This is not an academic question: Technical choices of network design made without considering legal, social,

economic, and administrative issues may prove to limit a new network's utility. Also, changes in law or administration needed to promote adoption may take years.

However appealing it may be technically, considering large scale modifications to the existing Internet raise a number of issues for attention:

- What can we learn from other instances of transition from one large-scale infrastructure to another? The history of technological innovation provides a rich and varied set of case studies of adoption, and failure thereof. The transition from the telegraph to the telephone took decades; the adoption of HDTV required a carefully crafted political process and important changes in spectrum policy; Sweden switched from driving on the left-hand side of the road to the right-hand side in a single day.
- What are the possible transition paths to a "clean slate" network? These could, for example, range from a DoD-only network to a process of widespread global adoption, with the current Internet being rapidly phased out.
- How do the technical characteristics of "clean slate" networks match up with the suite of possible adoption paths? Identifying potential gaps can shape future research directions for a "clean slate" network.
- How do existing administrative structures and economic forces match with potential trajectories for "clean slate" network adoption? This question takes on even more relevance because of ongoing discussion of Internet governance at the World Summit on the Information Society.

This discussion is not to suggest that widespread modifications to the existing Internet are inappropriate, only that the challenges of widespread modification are not just technical, but involve a complex set of social, economic, legal, and administrative issues that are poorly understood.

5.2 Social acceptance through incentives

Even given the many technical, legal, and diplomatic challenges to cyber attribution, achieving social acceptance is, perhaps, the greatest challenge to attribution. An Internet search for "Internet privacy" (quotes included) results in about 3,970,000 hits, while a search for "Internet attribution" (quotes included) results in 166 hits. The top hits on "Internet privacy" include a Wikipedia entry describing risks to privacy, Internet privacy resources, protecting privacy on the Internet, Internet Privacy Coalition pages, and Internet privacy law. The top hits on "Internet attribution" include a redirect to an unrelated service provider, an individual's blog, and a message board dealing with digital rights management (DRM) topics. Although these searches do not constitute a scientific study, they are somewhat representative of the Internet community. There is very little social acceptance for attribution because the average Internet user does not perceive attribution as providing positive value.

Rational parties behave according to incentives. Today's Internet lacks an incentive structure to encourage positive forms of attribution and discourage negative forms of attribution. Positive forms of attribution include methods to identify parties responsible for malicious behavior both during and after the act. Negative forms of attribution include attempts to discover the identity of those engaged in non-

malicious exchange of ideas. Today's Internet has very few incentives to discourage crime and malicious behavior: the rewards for malicious behavior are high, while the risk of being caught is low. Neither are there incentives for users to adopt attribution, a necessary ingredient to shift the reward structure in favor of non-malicious activities. The value of attribution is not understood by an average user. Most individuals are told that non-attribution (anonymity, privacy, and repudiation) is the essence of the Internet. It is certainly key to free exchange of ideas, as well as to malicious behavior.

Attribution can be used as a tool to create both positive and negative incentives, increasing value for the average user and promoting growth of individual, business, and governmental uses:

- Individuals should find value in reliable financial and business transactions and develop confidence that fraud and theft can be detected, attributed, and prosecuted.
- Businesses should have meaningful policy, legal, and technical methods to manage the risk associated with malicious Internet activity that disrupts business interests. One metric of success might be standard insurance models that protect against Internet disruptions. This would indicate relatively mature risk models and allow businesses to better manage risk.
- Governments are increasingly reliant on information technology (IT) and the Internet for routine and emergency operations. For many governments, Internet technologies serve as the backbone for intra-governmental communications, financial transactions, policy announcements, public relations, and emergency operations.

The importance of the global information infrastructure dictates the need to develop incentives for parties to adopt value-adding attribution methods. More important, we need to begin communicating the value of attribution so as to gain widespread understanding and acceptance.

5.3 A technical approach to attribution incentives

The key to creating positive incentives for Internet users to adopt, and eventually demand, attribution of online actions is identifying a set of online activities in which something of value to the user is at risk. Examples of such activities might include: online banking, online tax preparation, online medical services, and personal information management. To illustrate one possible approach, we will use online banking as an example.

It is possible to construct a logical overlay on the Internet that makes attribution possible among a set of online banking customers. That is, all relevant actions taken by any of the users electing to join the attribution overlay can be fully attributed to that user. This provides an online environment in which users are increasingly accountable for their actions. For applications in which transaction accuracy is expected, well-intentioned users will find value in a system in which online actions can be attributed to a valid user. Clearly, malicious actions taken by a non-member of the attribution overlay cannot necessarily be attributed to the offending party, but the attribution overlay can facilitate an investigation by proving that the action was generated outside of the membership community, protecting the user from fraudulent transactions. Also, if an attributed action is repudiated by a user of the attribution overlay, that user's computer can be segregated from the attribution overlay and examined for compromise.

The attribution overlay provides users with greater transactional accuracy, protection from Internet attacks originating outside the attribution overlay, and protection of machines inside the attribution overlay. The attribution overlay concept does not protect against all forms of exploitation, but represents a step in the right direction. As is typical with security services, the implementation will inevitably introduce unanticipated complexity and vulnerability, and will therefore require continuous refinement.

One possible implementation of an attribution overlay is based on a root of trust made available to individual users by an attribution group management process. The trust root is responsible for authenticating the origin of inbound messages and sealing attribution data of outbound messages through a digital signature technique. Attribution data must include the user, the machine, the message content and context, and the intended destination. The attribution data might also include geo-location, source routing data, required responses, and a two-way secure handshake. To avoid simple system compromise, unique keys used for authenticating and sealing attribution data must be well-protected in a hardware device, such as a trusted platform module (TPM) or a smart card, which may serve as the root of trust.

The purpose of this approach is to make it more difficult for a malicious actor to exploit the system exclusively using software. Outbound messages that require attribution can be presented to the TPM or smart card to verify message format and complete the digital signature. Inbound messages can be presented to the TPM or smart card to confirm origin, allowing each system an opportunity to ensure that all messages belong to the attribution overlay network before processing the messages. Note that this implementation will allow participants in the attribution overlay to restrict message processing to only those messages that originate within the attribution overlay. Further, it will allow operational security personnel to determine whether a compromise originated from within the attribution overlay or from an outside source. This simple distinction can help protect members of the attribution overlay from certain classes of fraud.

In addition to using a root of trust at each end point, there may be value to incorporating trust at various points within the network. These additional points of trust can attest, through a similar attribution process, to the existence of messages in the network at various times, providing investigators with the right information and strong confidence in that information. This would be similar to the process employed by physical investigators, who often rely on recording devices: automatic teller machine (ATM) cameras, traffic cameras, electronic toll collectors, credit card records, and dynamic host configuration protocol (DHCP) data.

Any attribution overlay must have a process to enroll and revoke users. This is a very difficult problem that is currently being studied by digital identity management experts. There are many approximate solutions for managing public key infrastructures that can serve as the basis for an attribution overlay. As new methods for identity management emerge, the attribution overlay can take advantage of them. Although the concept of an attribution overlay shares many challenges with digital identity management, the attribution overlay has one major advantage: uniform control. That is, the attribution overlay can be constructed by a single entity, such as a bank, and managed according to policy fully controlled by that entity. Under this model, many banks and online service providers can construct logical attribution overlays. Users can elect to join multiple attribution overlays,

making all actions relevant to that service provider attributable to the individual user. Note that it is also possible to use overlay technology to construct a non-attributable network for freely exchanging ideas.

Traditional attempts to create a network overlay have required that all users fully participate in the network overlay. The approach suggested above does not require full participation and is intended to gain steady acceptance for attribution services by communicating and providing value to a specific community. For example, an online financial service provider might offer value to customers electing to join the attribution network in the form of account insurance.

5.4 Reducing the malicious noise floor

Consider a rough classification of malicious activity that consists of three categories: nuisance, crime, and politically-motivated (terrorist and nation-state). The vast majority of reported malicious activity falls in the first two categories, and organizations such as the US Computer Emergency Response Team (CERT) devote considerable resources to tracking the frequency and type of such malicious activity. This is possible because many malicious methods used by parties to conduct low-level crime and nuisance are well-understood by the computer security community. Because attribution is not currently possible, our approach is to measure and track these types of malicious activities. The effort we expend to track low-level crime and nuisance activities detracts from more important security activities, namely high-level crime and politically-motivated activities, such as terrorism. Therefore, low-level malicious activity has two undesirable results: First, it consumes valuable resources; second, it raises the amount of malicious noise, making it more difficult to detect more damaging forms of malicious activity.

It is possible to discourage low-level malicious activity using low-grade attribution. It is not necessary to prosecute every low-level crime or resolve every low-level malicious activity to the location, computer, and individual. It is necessary to increase the probability that an individual engaging in malicious activity can be identified, prosecuted, and punished. Many interstate drivers know which states tolerate speeding and which do not. They know when the risk of being caught is high and when the penalty is steep. As a result, they tend to slow down in those states. An example of an Internet "speed trap" is the Record Industry Association of America (RIAA) campaign to identify violations of copyright through Internet file-sharing.

One technical approach to reduce the malicious noise floor might be to implement simple technical features that make it easier for cooperating Internet service providers to trace individual packets and flows, making it possible to issue the equivalent of "Internet traffic tickets." The legal model of traffic ordinances, which serve as penalties for violating rules of proper conduct, may extend into Internet activities. As with other incentive systems described in this paper, attribution is key to enabling such technology.

5.5 Supporting Legal Development

Western legal systems are complex and ever-evolving. Public law is based on proscription of actions deemed objectionable by society while private law is based on relationships between individuals. Identity is fundamental to both forms of law. To better understand the role of Internet attribution as a key enabler to support legal evolution, consider two legal scenarios: First, consider a criminal using the Internet to break existing laws, such as fraud. It is illegal to engage in deliberate

misrepresentation that causes another party to suffer damages, irrespective of the method. In this case, the role of the Internet is to facilitate the criminal activity. If a criminal elects to use the Internet in order to be more criminally productive, existing law uses a counting system to increase penalties. If the criminal elects to use the Internet to avoid detection, existing law is irrelevant. Therefore, in this example, detection of criminal activity and attribution to the offending party are critical to allowing the legal system to work. Note that although there can be additional laws proscribing the use of the Internet to commit fraud, they are not necessarily required in this case. What is required is a high likelihood that detected criminal activity can be attributed to the offending party.

In a second scenario, consider a party engaging in malicious activities that are purely cyber in nature, perhaps indiscriminate port scanning or denial of service. For newer crimes, clear legal guidance may not yet exist. Moreover, it is not possible to establish precedent without a specific case involving an alleged offense resulting in damage. In this scenario, Internet attribution is critical to identifying the offending party and pursuing legal action, which may include new legislation. Internet attribution is necessary to identify offending parties whenever the Internet is used to facilitate crime.

5.5.1 Internet attribution is only part of the picture

There are types of crime that involve the Internet, but are broader than the Internet, such that a lack of attributable actions on the Internet prevents sufficient overall attribution of the offending party. In such cases, Internet attribution is required as part of a larger attribution effort, rather than a standalone solution. Traditional law enforcement and business investigation techniques must evolve to include Internet attribution techniques and technologies.

Previously we suggested an attribution overlay as a method to create incentives for individuals to adopt attribution. Although this model may reduce fraud for an Internet business, the business case is somewhat lacking, since the business shifts risk from customers to itself by offering the attribution overlay. The value to the business is realized through enhanced investigative tools made possible by the attribution overlay. This technology can identify, with unprecedented confidence, when suspected fraudulent activity originated with a customer and when it came from an outside source. This data, and attribution function, can be used by law enforcement officials to begin a more comprehensive investigation. As comprehensive investigations involving both the cyber and physical domains evolve, businesses will benefit from the deterrence created by prosecuted cases and the increasing risk to mid- and high-level criminals. Note that a similar, possibly identical, comprehensive attribution function must be provided to advance private law as well, since contract disputes and torts will likely involve the Internet, but, as with criminal law, be broader than the Internet alone.

5.6 *The need for improved international cooperation*

5.6.1 Cooperation at a deeply technical level⁵⁴

With many cyber attacks crossing multiple jurisdictions, and with the increasing need for rapid and accurate attribution capabilities, there is a need for technical

⁵⁴ This section is drawn from Lipson, pp. 47-48.

cooperation far exceeding the agreements in principle now extant. Such cooperation may in fact be nascent in the developing system of CERTs, and, while directly of benefit in providing for improved attribution, would be of great value for cyber security in general.

Such a multilateral technical research, engineering, and advisory capability would fill some important gaps by:

- Researching and recommending the best attribution techniques
- Providing on-going support for a multilateral attribution capability
- Providing ongoing training and awareness for cooperating incident response and investigatory teams world-wide
- Making recommendations to international engineering bodies (e.g., the IETF) for protocol improvements and standards creation in support of member state requirements for tracking and tracing attackers
- Interacting with those creating cyber law and policy to ensure that the technical and non-technical approaches complement and support one another
- Helping ensure interoperability of the attribution infrastructures and technologies used by cooperating entities
- Assessing the results of cooperation already undertaken by technical and law enforcement agencies in order to provide feedback for continual improvement.

Ideally, such cooperation would involve sharing of information and cooperation about:

- Vulnerability information
- Incident data
- New methodologies and techniques for attribution and tracking, including both hardware and software tools
- Best practices
- Intelligence on the latest hacker capabilities and trends (including means of evading attempts to establish attribution).

Other desirable characteristics of a system of technical cooperation would include:

- Stability and continuity of the technical team to develop and maintain world-class expertise, particularly since technical information regarding attacker-defender capabilities and other technologies has a very short shelf-life. Informal cooperative arrangements will likely be inadequate in this regard.
- A global incident response capability (a multilateral incident response team). Such a team would be fully involved in day-to-day incident response operations, possibly raising issues of jurisdiction or control from the perspective of individual nation-states or other participating entities.

5.6.2 A common multilateral policy framework⁵⁵

There is need to formalize the cooperation and collaboration necessary to provide attribution across administrative, jurisdictional, and national boundaries. Such a formalized framework should consider:

⁵⁵ For a fuller discussion see Lipson PP. 49-53 and Wheeler, pp. 43-52.

- The organizations covered by the framework. It may be natural to consider policy frameworks as being among nation-states, but issues of attribution involve the interests of corporations and other organizations as well as nation-states.
- The amount of information to be collected. Technological issues may limit the amount of information collected to assist in attribution, but policy choices need to be made clear as far as under what circumstances tracking data can be collected, retained, and used.
- The amount of information to be shared, including the speed at which information will be shared, and the equipment and processes to be employed. This will become increasingly important as processes for attribution are automated. Policy choices as to the extent of information sharing need to be incorporated into the technical approaches. Equally, the ephemeral nature of evidence in cyber space makes speed in response essential; hence an effective policy framework will have to incorporate elements of a multilateral technical assistance function.
- Period of retention for data that may be used for tracking evidence of a crime, treaty violation, or other misdeed. There has, for example, been considerable concern that the Convention on Cybercrime defines too broadly the tracking data to be collected by ISPs, thereby threatening privacy.
- Types of incidents for which information will be shared; for example, what about an instance (which occurred during the release of the "I Love You" virus) in which the act is a crime in a victimized country, but not in the country of origin?
- Extra-territorial evidentiary seizure. One can envision an agreement that permits electronic extra-territorial seizure of data necessary to establish attribution, but under explicitly defined circumstances and with protections that will help prevent abuse. These might include placing evidence in escrow until proper legal authorities are provided for its access; if these are not provided, then the evidence can be destroyed or access withheld.
- Appropriate range of responses to an attack once attribution has been established. While this issue may appear to be outside the scope of the attribution issue, appropriate attribution, as previously discussed, depends on the possible range of responses. Conflicts might arise if, for example, party A might use attribution evidence as the basis for actions that party B, which has information essential to the attribution, considers inappropriate.
Range of response also needs to consider the time-frame. Extra-territorial defensive or retaliatory actions by the victim of an electronic attack might be justified (by, among other standings, Article 51 of the UN charter, which confirms a nation's inherent right of self-defense) while a cyber-attack is underway. It is less clear what responses are appropriate once an attack appears to be over; the severity of the attack, the likelihood of further attacks, and other considerations all come to bear. These are issues for which a pre-existing framework would be of value.
- Cost-sharing arrangements. Maintaining capabilities for proper attribution is not costless, nor are the special actions required to provide attribution for a specific event.

- Procedures for dealing with non-participants in the policy framework, and with untrusted jurisdictions. The Convention on Cyber crime is not a global agreement, nor is it likely that any policy framework for attribution will include all relevant entities. Non-participation or un-trusted entities obviously make attribution more difficult; it may be that there needs to be an agreed upon process for arriving at mutually agreed upon (by participants in the proposed policy framework) understandings as to what constitutes adequate attribution in the face of information from un-trusted sources.
- Adjudication in the event of mis-attribution. Mistakes may occur; actions in response to cyber attacks attributed to a particular source may subsequently prove to be based on incorrect assessments. An established process for addressing this issue is needed; there are a range of international procedures for dispute resolution, so that it may not be necessary to create new institutions. The key point is that these procedures should be worked out in advance.

6 Conclusion

Attribution is necessary to adequately address malicious Internet activity, irrespective of the nature of that activity. The required scope and confidence of attribution depends on the severity of malicious activity and the parties involved or impacted. Sufficient attribution can be achieved through a series of gradual steps, but cannot be achieved by simply introducing new technology or policy. Attribution requires a system of acceptance, cooperation, technology, and traditional investigation supported by policy, law, and treaty. Designing such a system is complicated by indecision about what constitutes malicious Internet activity, the tension between necessary attribution and necessary non-attribution, and the fact that the entire international must participate in any useful attribution implementation.

One key step to developing a system of attribution is gaining user acceptance. Well intentioned parties must be encouraged, through incentives, to accept and value attribution, lest they work around any system we might create. In order for a system of attribution to work, good citizenship must be the norm, and an incentive structure must reinforce desirable behavior. While attribution is necessary to discourage and punish malicious behavior, selective non-attribution is a critical feature of the Internet and must be preserved, and even strengthened, so as to facilitate free exchange of ideas and protect individuals from oppressive regimes.

As developed countries have become highly reliant on the Internet for both business and government operations, Internet infrastructure has become a target of military-style attack. Perhaps the greatest challenge for an attribution system will be identifying nation-states that elect to engage in war-like cyber activities, as these untrustworthy parties have no interest in supporting effective attribution. This challenge can only be overcome through a comprehensive approach, such as the system we adopted for global nuclear non-proliferation, which developed multilateral technical and non-technical approaches in a mutually distrusting environment.

Attribution is a complex problem and structuring investment will be a significant challenge. Designing and implementing a working attribution system will require continuous refinement, balancing many social, political, and technical requirements.

Metrics of progress may include: the emergence of an Internet disruption insurance market for online businesses, meaningful definitions of malicious online activity, technically informed law, significant reductions in low-level malicious activities, greater specialization of operational computer security staff, and the emergence of attribution service businesses.

7 Acronyms

US	United States
IP	Internet Protocol
IETF	Internet Engineering Task Force
ITU	International Telecommunications Union
VPN	Virtual Private Network
IDS	Intrusion Detection System
DOS	Denial of Service
DDOS	Distributed Denial of Service
NATO	North Atlantic Treaty Organization
UN	United Nations
PAN	Perfect Attribution Network
NAA	Non-Attribution Application
NAAO	Non-Attribution Application Overlay
DNS	Domain Name System
DNSSec	Domain Name System Security Extensions
BGP	Border Gateway Protocol
RFC	Request for Comments
ISP	Internet Service Provider
FISA	Foreign Intelligence Surveillance Act
FBI	Federal Bureau of Investigation
FATF	Financial Action Task Force
ICMP	Internet Control Message Protocol
NSF	National Science Foundation
FIND	Future Internet Network Design
DoD	Department of Defense
GIG	Global Information Grid
DRM	Digital Rights Management
IT	Information Technology
TPM	Trusted Platform Module
ATM	Automatic Teller Machine
DHCP	Dynamic Host Configuration Protocol
CERT	Computer Emergency Response Team
RIAA	Recording Industry Association of America