Chapter 7

ATTRIBUTION OF CYBER ATTACKS ON PROCESS CONTROL SYSTEMS

Jeffrey Hunker, Robert Hutchinson and Jonathan Margulies

Abstract The attribution of cyber attacks is an important problem. Attribution gives critical infrastructure asset owners and operators legal recourse in the event of attacks and deters potential attacks. This paper discusses attribution techniques along with the associated legal and technical challenges. It presents a proposal for a voluntary network of attributable activity, an important first step towards a more complete attribution methodology for the control systems community.

Keywords: Process control systems, cyber attacks, attack attribution

1. Introduction

United States Presidential Decision Directive NSC-63 (PDD 63) [1] listed the infrastructures that are critical to national security. The directive also stressed the need for public-private partnerships to identify and mitigate critical infrastructure (CI) vulnerabilities. Assessment and modeling efforts resulting from the recognition of vulnerability as outlined in PDD 63 have revealed significant interdependencies between infrastructures [4]. These efforts have also shown that CI protection is a global problem and that the global infrastructure depends on the proper operation of standardized, as well as specialized, information technology.

All the CIs are supported to varying degrees by the global information infrastructure, much of which is built on commodity technologies such as the TCP/IP protocol suite and backbone networks. Many CI assets, such as those responsible for energy production and distribution, require specialized control systems for safe and reliable operation. The growing convergence of specialized process control systems with general information and communication technologies (ICTs) is exposing control systems and the CI assets they manage to common operating system and Internet threats [5].

Attack attribution can provide new types of protection for CI assets. It gives asset owners and operators legal recourse in the event of attacks and deters malicious activities. The development and application of attribution techniques can drive law and policy, create incentives for cyber security, reduce threats and manage risk.

2. Attack Attribution

Attribution is the determination of the identity or location of an attacker or an attacker's intermediary [6]. This paper focuses on attribution techniques for attacks launched over computer networks. Attribution is also important for physical attacks and social engineering attacks; however, these attacks are outside the scope of this paper. For the purposes of this paper, we consider attribution to be the identification of intermediaries who may or may not be willing participants in an attack. Note that determining motivation, particularly by technical means, is difficult at best. This problem is even more challenging when applied to intermediaries.

2.1 Importance of Attribution

The ability to identify the source of a cyber attack is the basis for taking action against the perpetrator. Legal and policy frameworks for responding to cyber attacks cannot work unless there is adequate attribution; these frameworks remain incomplete because there is insufficient basis (attribution) to actually use them. Attribution helps create a system of deterrence. Without the fear of being caught, convicted and punished, individuals and organizations will continue to use the Internet to conduct malicious activities.

Attribution also offers other benefits. Information gained during the process of attribution can be used to improve defensive techniques. Even partial attribution can provide the basis for interrupting attacks in progress and defending against future attacks and mitigating their effects.

While attribution is important, non-attribution can be just as vital to protecting radical ideas and minority views in oppressive regimes. The Internet has become a powerful medium for sharing opinions. For many users, the anonymity provided by the Internet – and the consequent freedom from retribution – makes it one of the only methods for freely expressing ideas. Mechanisms developed to facilitate attribution must enforce non-attribution for the purposes of sharing opinions and ideas. A well-crafted attribution mechanism should identify entities who engage in malicious behavior as defined by laws and/or policy. However, the mechanism should also make it impossible to attribute freely exchanged ideas, especially for purposes of retribution.

2.2 Attribution in Process Control Systems

Critical infrastructure owners and operators rely on standard ICTs to monitor and control physical processes. The adoption of these technologies in the

global energy, chemical, transportation and service infrastructures injects ICT vulnerabilities into formerly stand-alone process control systems. This exposes CI assets to a slew of attacks by external entities, including exploits launched over the Internet. Infrastructure owners and operators are struggling to understand the new risks and incorporate appropriate risk mitigation techniques. ICT-based systems primarily rely on fortification mechanisms (e.g., firewalls, access control lists and physical access controls) to mitigate risk. Applying ICT security mechanisms and risk models to CI control systems is inadequate due to the major differences that exist between CI and ICT assets in terms of their physical layout, operating environments, security goals and threat space.

For reasons of complexity and cost, CI asset owners and operators have outsourced significant aspects of their operations. Control system hardware and software manufactured globally are customized to a particular operation, effectively incorporating vendors into the operational lifecycle. Control system vendors frequently provide onsite support and are often granted remote access to monitor system status and perform routine administration tasks. Given the number of entities – employees, equipment manufacturers, consultants and employees from partnering companies – that access CI assets, threats are better modeled as originating from malicious insiders than external attackers. Even so, CI assets are protected by fortification, a strategy that ignores insiders, which are a system's greatest threat.

Cyber attribution can help protect the essential command, control and communication functions of modern infrastructures. Attribution makes it possible to enforce existing laws and treaties and fine-tune law and policy to better protect CI assets. Coupled with the ability to detect malicious cyber activity, attribution and the prospect of penalties deter attackers, thereby reducing threats. Technology is a necessary component of cyber attribution, but it is insufficient by itself. To be successful and lasting, an attribution strategy must acknowledge the need for local and international cooperation while recognizing that the technical component must evolve with advances in technology.

2.3 Ideal Attribution

To provide a context for our work, it is necessary to define the ideal attribution system to which we aspire. As international law, policy and technology evolve, this definition must evolve as well.

Attribution should exist in a global context with overt support from all nation states and organized entities. Sufficient attribution of malicious cyber activity should be possible even when some entities do not cooperate fully. Attribution should only be sought for malicious activity.

An ideal attribution system should make it possible to detect all cyber attacks and determine the source and intent of each attack with sufficient precision. The results should be verifiable and specific enough to justify any response that has been agreed to in advance. The attribution system should generate threat data that informs the development and implementation of CI technologies and defensive strategies. Finally, the attribution system should

enable new international agreements and pave the way toward comprehensive risk management.

3. Difficulty of Attribution

The Internet's architecture and its evolving administrative and governance systems make the attribution of cyber attacks extremely challenging.

3.1 Internet Architecture

The Internet has no standard provisions for tracking or tracing. A sophisticated user can modify information in IP packets and, in particular, forge the source addresses of packets (which is very simple for one-way communication). Attacks often employ a series of stepping stones where compromised intermediate hosts are used to launder malicious packets. Packets can also be changed at hops between hosts; thus, attempting a traceback by correlating similar packets is ineffective when sophisticated attackers are involved.

3.2 Administrative Issues

Internet attacks exploit an administrative system that was established when the Internet community was, in essence, a trusted commune, not the current global virtual city with consequent malefactors and little shared sense of community.

Internet 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) does not provide a global system policy and technical framework that the International Telecommunications Union (ITU) provides for the telephone system. According to Lipson [3]:

> "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."

3.3 Limited Business Support

Businesses and government agencies routinely respond to malicious cyber activities by rebooting critical servers, restoring lost data and identifying and eliminating the vulnerabilities. The fundamental goal of IT professionals in industry and government is to maintain operations, and this is particularly true

with regard to control systems. Because attribution is so difficult, few organizations are interested in investigating malicious Internet activity and moving them through the legal system. The current protection model, therefore, is primarily aimed at building improved fortifications and attempting to withstand constant attempts to overcome the fortifications.

3.4 Technical Impediments

Several technical impediments limit the effective attribution of attacks:

- Tunneling impedes tracking, but it is also very useful for creating virtual private networks (VPNs) that are important for security.
- Hackers often destroy logs and other audit data once they gain system access.
- Anonymizing services are valuable to Internet users (e.g., to facilitate political discourse in countries with repressive regimes). While anonymizers can be defeated in theory, there are numerous practical difficulties to achieving attribution when a sophisticated user desires anonymity.
- Even if an attack packet can be attributed to an IP address of a host computer, it is difficult to link the IP address to the actual perpetrator. A perpetrator can decouple his physical identity from an IP address by using cyber cafes, public Internet facilities (e.g., libraries) and prepaid Internet address cards that can be purchased from service providers without any personal identification.
- Short storage intervals on routers, especially those located at or near the high-speed core of the Internet, require forensic techniques to be extremely rapid (i.e., capture evidence before the router cache is overwritten). Alternatively, new capabilities have to be created to proactively preserve routing information.
- Sophisticated attacks that are extremely fast or extremely slow (that may execute over a period of months) are difficult to detect.
- Attribution techniques themselves have to be secured against attacks and subversion. Software used for authentication and data used for attribution must be protected. Moreover, attribution techniques should not create additional avenues for exploitation (e.g., a new DOS attack against the system).

3.5 Liability for Attributable Activities

Liability for the causes and consequences of cyber attacks is an undeveloped field within U.S. jurisprudence. All of the questions related to liability for domestic-only cyber attacks are equally applicable to attacks that traverse international boundaries. However, there are no clear answers to cyber attack

liability questions, and relevant case law is sparse. This issue is of great concern to CI owners and operators because their losses due to cyber attacks can be extremely high.

Several important liability questions must be considered [3]: a major issue is the liability exposure for various entities – perpetrators; vendors whose software and/or hardware made the attack possible; owners, operators and administrators of intermediate (zombie or anonymizer) systems that participate in the attack or obscured the attack source; and service providers who did not block the attack when notified or did not help trace the attack in accordance with international agreements.

Another issue is whether certain kinds of waivers on liability exposure should be provided to entities who participate in tracking and tracing. Also, there is the issue of liability for the owners of systems that participate in attacks without the owners' knowledge, and entities that provide anonymizing services.

It is clear that liability will eventually form an important component in the policy framework for cyber security in general. This framework will affect the range of feasible options for effective attribution of cyber attacks.

3.6 Feasibility of Technical Solutions

Technology alone cannot provide attribution. To demonstrate this fact, we consider an ideal network with perfect attribution. 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 PAN foundation provides attribution services that 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 (e.g., process control system transaction management) 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 the 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 source and destination users by the underlying PAN, the attribution scope is limited to the immediate source and destination of each message.

One strategy for achieving non-attribution is to remove the point-to-point context from all messages by applying an NAA overlay (NAAO). As a simple example, consider an NAAO configured in a logical ring topology. Messages in the ring topology flow clockwise and each NAA instance receives all of its incoming messages from a single NAA instance (the instance that is immediately

counterclockwise) and transmits all outgoing messages to a different NAA instance (the instance that is immediately clockwise). The NAAO provides strong confidentiality of all messages, generates random messages to maintain roughly constant data flow bandwidth regardless of message content, and abstracts the actual message source and destination from the PAN. Direct PAN features cannot identify which messages are authentic, the real source and destination of messages or the real source of messages leaving the NAAO. Furthermore, constant bandwidth utilization in the ring topology makes traffic analysis very difficult (this is not a function of the PAN).

This simple example demonstrates that a purely technical attribution solution does not exist. It is possible to develop features that facilitate attribution such as unique communication keys, traceback and logging. However, these solutions may not provide sufficient attribution for most instances of malicious behavior. Essential to providing attribution services is a user base that is interested in supporting the services.

4. Limitations of Current Solutions

Table 1 summarizes the technical solutions for attribution that are currently being employed, developed or considered. Details and drawbacks of these solutions are discussed extensively in [2].

The principal drawback is the lack of widespread adoption of new technologies. The record of Internet-wide adoption of new technologies has historically been poor. This is largely due to the new capabilities required of routers, but the problem is of a more general nature. Specific examples include IPv6, DNS security (DNSSec) extensions, and modifications to the Border Gateway Protocol (BGP).

These cases have two common issues. The first is that changes are effective only if they are adopted uniformly across the community. The second is that the costs and burden accrue to the individual entity, while the benefits of the changes are distributed and system-wide. Unlike the telecommunications sector where the ITU has an effective mechanism for creating and enforcing technical requirements, IETF's request for comment (RFC) process is essentially voluntary. Hence, a major policy issue is how to create incentives or a regulatory system that would ensure that system-wide changes are implemented.

5. Attribution Steps

Several options are available for designing an attribution scheme.

- Design a scheme similar to that used in telephone systems where attribution is automatic – a call is traced back to the originating number. A user may turn this feature off, but this decision can be overridden by the courts.

- Design a system where users can opt in or opt out without any subsequent recovery ability by the system.

Table 1. Technical solutions for attack attribution.

Technique	Description
Hash-Based IP Traceback	Routers store hash values of network packets. Attribution is done by tracing back hash values across network routers.
Ingress Filtering	All messages entering a network are required to have a source address in a valid range; this limits the range of possible attack sources.
ICMP Return to Sender	All packets destined for the victim are rejected and returned 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.
Trace Packet Generation (e.g., iTrace)	A router sends an ICMP traceback message periodically (e.g., every 1 in 20,000 packets) to the same destination address as the sample packet. The destination (or designated monitor) collects and correlates tracking information.
Probabilistic Packet Marking	A router randomly determines whether it should embed message route data in a message; this routing data is used to determine routes.
Hackback	Querying functionality is implemented in a host without the permission of the owner. If an attacker controls the host, this may not alert the attacker; thus, the information is more reliable.
Honeypots	Decoy systems capture information about attackers that can be used for attribution.
Watermarking	Files are branded as belonging to their rightful owners.

- Design multiple networks, each providing a different level of attribution; some networks may provide no attribution. Users are free to choose the networks from which they accept packets.

5.1 Implementing Clean Slate Approaches

The Internet was not designed for the purposes for which it is now used, 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's Future Internet Network Design (FIND) and the Department of Defense's Global Information Grid (GIG). The goal of these projects is to develop new networks of networks that address current and future needs.

The challenge faced by these projects is that network architectures are not just about technology – they also involve social, legal, administrative and economic infrastructures. Secure BGP and IPv6 have faced slow acceptance as much for technical reasons as for economic and administrative reasons.

No plans exist for the adoption of a clean slate network. Indeed, if some of the current clean slate work is successful, it would – in colloquial terms – be like the dog that chases the car finally catches it! This is not an academic question: technical designs that do not consider legal, social, economic and administrative issues definitely limit the utility of the new networks. Also, legal and/or administrative changes needed to promote the adoption of a new network architecture may take years.

However appealing it may be technically, the consideration of large-scale modifications to the Internet raises several interesting questions:

- What can we learn from other instances of transition from one large-scale infrastructure to another? The history of technological innovations provides a rich and varied set of case studies of adoption and failures. The transition from the telegraph to the telephone took decades; the adoption of HDTV required a carefully-crafted political process and major 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 creating a DoD-only network to one that is adopted globally, with the current Internet 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 clean slate networks.

- How do existing administrative structures and economic forces match with potential trajectories for clean slate network adoption? This question takes on even more relevance given the discussion of Internet governance at the World Summit on the Information Society.

It is important to note that we do not believe that widespread modifications to the existing Internet are inappropriate. Instead, our position is that the challenges to any modification are not just technical in nature, but involve a complex set of social, economic, legal and administrative issues that are poorly understood.

5.2 Technical Approach for Limited Attribution

It is possible to construct a logical overlay on the Internet that makes attribution possible within a subset of users, in this case, process control system operators. All relevant actions taken by an entity electing to join the attribution overlay can be fully attributed to that entity. This provides an online environment in which members are increasingly accountable for their actions.

For control system applications – where transaction accuracy is critical – the value of attribution is obvious. The control systems environment provides a unique opportunity to develop and demonstrate a limited attribution service because the drawbacks of attribution are minimal while the benefits are high. The ability to attribute control system activities promotes integrity as well as trust in operational decisions. Attribution also makes it possible to discover and discourage attempts at compromising process control systems, even those perpetrated by insiders.

Clearly, malicious actions taken by a non-member of an attribution network 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 members from fraudulent transactions. Also, if an attributed action is repudiated by a member of the attribution overlay, that member's system can be segregated from the attribution overlay and examined for compromise.

The attribution overlay would provide control systems with greater transactional accuracy. Also, it would provide protection from Internet attacks originating outside the attribution overlay and protect machines inside the attribution overlay. The attribution overlay concept is not a defense against all forms of exploitation, but it is a step in the right direction. As is typical with security services, its implementation will inevitably introduce unanticipated complexities and vulnerabilities, and will, therefore, require continuous refinement.

One possible implementation of an attribution overlay is to use a root of trust made available to individual members 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 using a digital signature technique. Attribution data must include the member, machine, message content and context, and intended destination. The data may also include geo-location, source routing data, required responses and a two-way secure handshake. To avoid system compromise, the keys used for authenticating and sealing attribution data must be protected using a hardware device (e.g., trusted platform module (TPM) or smart card), which may serve as the root of trust.

This approach seeks to make it more difficult for malicious actors to exploit the system exclusively using software. Outbound messages that require attribution can be presented to the TPM or smart card to verify the message format and complete the digital signature. Inbound messages can be presented to the device 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 would allow participants in the attribution overlay to restrict message processing to those messages that originate within the attribution overlay. Also, it allows 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 types of fraud.

In addition to using a root of trust at each end point, there is value to incorporating trust at multiple points within the network. These additional points of trust can attest (via 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 the information. This would be similar to the process employed by physical investigators who often rely on recording devices: ATM cameras, traffic cameras, credit card records and DHCP data.

An attribution overlay must have a process to enroll and revoke members. This is a difficult problem that is currently being studied by digital identity management experts. Some approximate solutions for managing public key infrastructures can serve as the basis for an attribution overlay. However, as new methods for identity management emerge, the attribution overlay can be revised to take advantage of them. Although attribution overlay shares many challenges with digital identity management, attribution overlay has the advantage of uniform control. Specifically, attribution overlay can be constructed by a single entity (e.g., a control system working group) and managed according to a policy controlled by that entity. Under this model, control system operators and service providers can construct logical attribution overlays. Members may elect to join multiple attribution overlays, making all the actions relevant to that service provider attributable to the individual member. Note that it is also possible to use overlay technology to construct a non-attributable network for the free exchange of ideas.

Traditional attempts at creating network overlays require all the members to fully participate in the network overlay. Our approach does not require full participation and is intended to steadily gain acceptance for attribution services by providing value to the control systems community over time.

5.3 Reducing the Malicious Noise Floor

Malicious activity may be broadly divided into three categories: nuisance, criminal and politically-motivated (terrorist or nation-state) activities. The vast majority of reported activity falls in the first two categories; organizations such as the U.S. Computer Emergency Response Team (CERT) devote considerable resources to analyze malicious acts. This is possible because the methods used by actors to conduct nuisance activities and low-level crime 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 expended to track them detracts from the ability to focus on 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 difficult to detect more damaging types of activities.

Low-level malicious activity can be discouraged using low-grade attribution. It is not necessary to prosecute every low-level crime and resolve every low-level

malicious activity to the computer and individual. All that is required is to increase the probability that an individual engaging in malicious activity can be identified, prosecuted and punished. Many interstate drivers know which jurisdictions tolerate speeding and which do not. They know where the risk of being caught is high and where the penalties are steep, and they tend to slow down in those jurisdictions. 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 for reducing the malicious noise floor is to implement features that make it easier for cooperating Internet service providers to trace individual packets and flows, which would make it possible to issue the equivalent of "Internet traffic tickets." The legal model of traffic ordinances, which stipulate penalties for violating the rules of proper conduct, may, thus, be extended to Internet activities. But, of course, attribution is key to enabling such an approach.

6. Conclusions

Critical infrastructure protection is a global problem. The adoption of commodity information, computing and networking technologies in critical infrastructure assets increases the overall risk by growing the space of exploitable vulnerabilities and exposing systems to new threats. While new technologies should be developed to help reduce vulnerabilities, it is equally important to seek strategies that manage the threats. Cyber attribution supports threat management by identifying malicious actors and uncovering their motives and methods, in the process, informing vulnerability reduction efforts. Attribution gives critical infrastructure asset owners and operators legal recourse in the event of attacks and deters potential attacks. Technological advances alone will not solve the attribution problem. Effective and lasting attribution services require active user participation, and sustained legal and policy efforts.

Acknowledgements

This work was partially supported by the Institute for Information Infrastructure Protection (I3P) at Dartmouth College, Hanover, New Hampshire, under Award 2003-TK-TX-0003 from the U.S. Department of Homeland Security.

References

[1] W. Clinton, Presidential Decision Directive 63, The White House, Washington, DC (fas.org/irp/offdocs/pdd/pdd-63.htm), May 22, 1998.

[2] J. Hunker, R. Hutchinson and J. Margulies, Roles and Challenges for Sufficient Cyber Attack Attribution, Research Report, Institute for Information Infrastructure Protection, Dartmouth College, Hanover, New Hampshire, 2008.

[3] H. Lipson, Tracking and Tracing Cyber Attacks: Technical Challenges and Global Policy Issues, Special Report CMU/SEI-2002-SR-009, CERT Coordination Center, Software Engineering Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania, 2002.

[4] P. Pederson, D. Dudenhoeffer, S. Hartley and M. Permann, Critical Infrastructure Interdependency Modeling: A Survey of U.S. and International Research, Report No. INL/EXT-06-11464, Critical Infrastructure Protection Division, Idaho National Laboratory, Idaho Falls, Idaho, 2006.

[5] T. Samad, P. McLaughlin and J. Lu, System architecture for process automation: Review and trends, *Journal of Process Control*, vol. 17(3), pp. 191–201, 2007.

[6] D. Wheeler and G. Larson, Techniques for Cyber Attack Attribution, IDA Paper P-3792, Institute for Defense Analyses, Alexandria, Virginia, 2003.