Chapter 16
Indirect Attribution in Cyberspace

Robert Layton
Federation University, Australia

Paul A. Watters
Massey University, New Zealand

ABSTRACT

We are now in an era of cyberconflict, where nation states, in addition to private entities and individual actors, are attacking each other through Internet-based mechanisms. This incorporates cyberespionage, cybercrime, and malware attacks, with the end goal being intellectual property, state secrets, identity information, and monetary gain. Methods of deterring cybercrime ultimately require effective attribution; otherwise, the threat of consequences for malicious online behaviour will be diminished. This chapter reviews the state of the art in attribution in cyberspace, arguing that due to increases in the technical capability of the most recent advances in cyberconflict, models of attribution using network traceback and explicit identifiers (i.e. direct models) are insufficient build trustworthy models. The main cause of this is the ability of adversaries to obfuscate information and anonymise their attacks from direct attribution. Indirect models, in which models of attacks are built based on feature types and not explicit features, are more difficult to obfuscate and can lead to more reliable methods. There are some issues to overcome with indirect models, such as the complexity of models and the variations in effectiveness, which present an interesting and active field of research.

INTRODUCTION

In 2012, U.S. President Obama officially recognised that the Stuxnet virus, which targeted SCADA controllers operating Iranian nuclear facilities, was a state based attack that originated from the USA and Israel (Sanger, 2012). In that recognition, the world moved towards an era where state sponsored cyberconflict is no longer a conspiracy theory (or probable scenario of the world), but an accepted fact. Recent reports by industry and ex-government officials have pointed to other countries like China also being responsible for other attacks, with one allegation being the theft of confidential trading information that led to millions in losses in negotiation potential (Fowler & Cronau 2013). Both the US and China are organising a treaty on "cyber-arms" (Arimatsu, 2012), with a view to recognizing acceptable limits on this fifth domain of war (the first four being

DOI: 10.4018/978-1-4666-6324-4.ch016

land, sea, air and space). However a fundamental component to the enforcement and effectiveness of such a treaty is missing. Without the adequate attribution of cyberattacks, treaties are worth little at best and can be used for the deliberate misdirection of blame at worse (Watters et al. 2013).

In the rush to uptake technology as a core component of critical infrastructure, nations have now found that many of the systems they rely upon are open to potential attack. This includes water systems, intelligence networks and trading information. To protect this critical infrastructure, investment into defences against cyberattacks has increased dramatically over recent years. Governments across the world are increasing their capability and capacity in both defensive and offensive cyber-based programs. While offensive capabilities are increasing, deterrence of cyberattacks has not caught up, as Guitton (2013) notes: "if the adversary knows that the likeliness for a threat of retaliation is low due to the uncertainty of attribution, deterrence is unlikely to function" (p96).

Attribution can be absolute, in that it identifies an actor responsible for a given attack, or relative in that it can tell us that two attacks have the same origin. As noted by Sigholm and Bang (2013) of cyberattack attribution, "the process of attaining positive attribution is perceived as being ineffective" (p. 167). A number of reasons are cited for this, including a lack of access to data, but also the lack of a process that facilitates effective attribution in cyberattacks. In cases where data is available, Sigholm and Bang notes that "the inability to define adequate filters, to make sense of the collected data, and to understand what is important and not, that constitutes the main problem (of attribution of cyberattacks)" (p. 167).

In most recent cases where a cyberattack has been attributed, there has often been a critical mistake on the part of the attacker. In a recent Mandiant APT1 report, the attackers left their name within the attacking programs, linking their attacks to a long online history (Mandiant, 2013). Such mistakes cannot be relied upon, nor expected to be uncovered in a timely fashion to determine if a country is breaking a treaty through a cyberattack.

Such mistakes, where they exist and the information can be trusted to be accurate, are highly effective pieces of evidence. One example is the use of an atypical and consistent misspelling by an author, which is one of the most effective forms of attribution for a written document (Juola, 2006). However such mistakes cannot be relied upon to exist. Therefore, they cannot form the basis of an effective attribution strategy that needs to be robust, trusted and timely to be used effectively. In addition, relying on commonly known features may open the risk of the attacker inserting deliberately misleading evidence to cause attribution to another actor. Modelling the vector of attack, content of the attack and other meta-data may not be as conclusive as a significant error on behalf of the attacker, but can be applied in more cases.

Cyberattacks, particularly at a state level, are generally technically capable, well resourced, and benefit from adding misdirection and complexity into their attacks. We argue that confidence in attribution can only come from indirect models, because direct features, such as the tracing of an attack through the network path it took, can be easily faked in complicated attacks. Indirect models aim to model intuitive and subconscious aspects to attacks which are more difficult to hide, as the user may not be aware of them. While sophistication is not a mutually exclusive condition of state-based attacks (Guitton & Korzak, 2013), this chapter focuses on such complex attacks.

It is often important to define exactly what is meant by a term, particularly when dealing with criminal behaviour, which often varies in terms of legality and definition in different contexts. For the purposes of this chapter, a computer or person is a *target* if there is some other actor wishing to attack them. We define the *victim* of an attack as someone who has been attacked, and the *attacker* as the person who initiated the attack. We also take the viewpoint as the defender of the victim or target, and therefore an *adversary* is someone