

# An Argumentation-Based Reasoner to Assist Digital Investigation and Attribution of Cyber-Attacks

Erisa Karafili

*University of Southampton*

June 3, 2020  
DFRWS EU

Erisa Karafili, Linna Wang, Emil C. Lupu



Funded by the EU's Horizon 2020 under the Marie Skłodowska-Curie grant agreement No 746667.

# Agenda

- 1 Introduction
- 2 An Argumentation-Based Reasoner
- 3 Conclusions and Future Work

The growing of connectivity increases the **security** challenges and the need for **efficient** countermeasures

Analyzing and attributing cyber-attacks permits efficient **attacker-oriented** countermeasures

- **Digital Forensics** techniques help the analysis and attribution
- These techniques suffer from the **quantity** and **quality** problem

# The Problem

## Problem

*The attribution process is a **difficult** one and there is a **need** to provide **help** to the analyst during this process*

- Attribution is mainly **human** based
- It suffers from human **errors** and is easily **biased**
- Explanations on the provided results are missing

# The Proposed Solutions

## Solution

*An automatic **reasoner** that helps the analyst to **analyze** the pieces of evidence and **attribute** the attack*

- Our solution **reduces** the human errors and bias
- It permits to work with **incomplete** and **conflicting** evidence
- It provides an **explainable** attribution

# An Argumentation-Based Reasoner

# An Argumentation-Based Solution

## Solution

An *automatic reasoner* (ABR) that helps the forensics analyst during the analysis and attribution process.

- ABR is based on **argumentation** and **abductive** reasoning
- It works with **incomplete** and **conflicting** pieces of data
- ABR works with **technical** and **social** evidence

# Preference-Based Argumentation Framework

Our solution uses a **preference-based argumentation** framework

## Definition

An *argumentation theory* is a pair  $(\mathcal{T}, \mathcal{P})$  of argument rules  $\mathcal{T}$  and preference rules  $\mathcal{P}$ .

The **argument rules**  $\mathcal{T}$  are a set of labelled formulas of the form:

$$rule_i : L \leftarrow L_1, \dots, L_n.$$

The **preference rules** are a set of labelled formulas of the form:

$$p : rule_1 > rule_2$$

where  $rule_1, rule_2$  are labels of rules in  $\mathcal{T}$ , and  $>$  is **higher priority relation** between the rules.

# A Simple Example

Given the argument pair  $(T, P)$  :

$$T = \{r_1 : \text{attackOrig}(X, \text{Attack}) \leftarrow \text{ipGeoloc}(X, IP), \text{attackSourceIP}(IP, \text{Attack}). \\ r_2 : \neg \text{attackOrig}(X, \text{Attack}) \leftarrow \text{ipGeoloc}(X, IP), \text{attackSourceIP}(IP, \text{Attack}), \\ \text{spoofedIP}(IP).\}$$

$$P = \{p_1 : r_2 > r_1\}$$

# A Simple Example

Given the argument pair  $(T, P)$  :

$$T = \{r_1 : \text{attackOrig}(X, \text{Attack}) \leftarrow \text{ipGeoloc}(X, IP), \text{attackSourceIP}(IP, \text{Attack}). \\ r_2 : \neg \text{attackOrig}(X, \text{Attack}) \leftarrow \text{ipGeoloc}(X, IP), \text{attackSourceIP}(IP, \text{Attack}), \\ \text{spoofedIP}(IP).\}$$

$$P = \{p_1 : r_2 > r_1\}$$

and the following evidence:

$$E = \{\text{attackSourceIP}(ip00, A_1), \text{ipGeoloc}(\text{countryC}, ip00)\}$$

# A Simple Example

Given the argument pair  $(T, P)$  :

$$T = \{r_1 : \text{attackOrig}(X, \text{Attack}) \leftarrow \text{ipGeoloc}(X, IP), \text{attackSourceIP}(IP, \text{Attack}). \\ r_2 : \neg \text{attackOrig}(X, \text{Attack}) \leftarrow \text{ipGeoloc}(X, IP), \text{attackSourceIP}(IP, \text{Attack}), \\ \text{spoofedIP}(IP).\}$$

$$P = \{p_1 : r_2 > r_1\}$$

and the following evidence:

$$E = \{\text{attackSourceIP}(\text{ip00}, A_1), \text{ipGeoloc}(\text{countryC}, \text{ip00})\}$$

the conclusion is:

$$\text{attackOrig}(\text{countryC}, A_1).$$

# A Simple Example

Given the argument pair  $(T, P)$  :

$$T = \{r_1 : \text{attackOrig}(X, \text{Attack}) \leftarrow \text{ipGeoloc}(X, IP), \text{attackSourceIP}(IP, \text{Attack}). \\ r_2 : \neg \text{attackOrig}(X, \text{Attack}) \leftarrow \text{ipGeoloc}(X, IP), \text{attackSourceIP}(IP, \text{Attack}), \\ \text{spoofedIP}(IP).\}$$

$$P = \{p_1 : r_2 > r_1\}$$

and the following evidence:

$$E = \{\text{attackSourceIP}(ip00, A_1), \text{ipGeoloc}(\text{countryC}, ip00)\}$$

the conclusion is:

$$\text{attackOrig}(\text{countryC}, A_1).$$

If the evidence is:

$$E = \{\text{attackSourceIP}(ip00, A_1), \text{ipGeoloc}(\text{countryC}, ip00), \text{spoofedIP}(ip00)\}$$

# A Simple Example

Given the argument pair  $(T, P)$  :

$$T = \{r_1 : \text{attackOrig}(X, \text{Attack}) \leftarrow \text{ipGeoloc}(X, IP), \text{attackSourceIP}(IP, \text{Attack}). \\ r_2 : \neg \text{attackOrig}(X, \text{Attack}) \leftarrow \text{ipGeoloc}(X, IP), \text{attackSourceIP}(IP, \text{Attack}), \\ \text{spoofedIP}(IP).\}$$

$$P = \{p_1 : r_2 > r_1\}$$

and the following evidence:

$$E = \{\text{attackSourceIP}(ip00, A_1), \text{ipGeoloc}(\text{countryC}, ip00)\}$$

the conclusion is:

$$\text{attackOrig}(\text{countryC}, A_1).$$

If the evidence is:

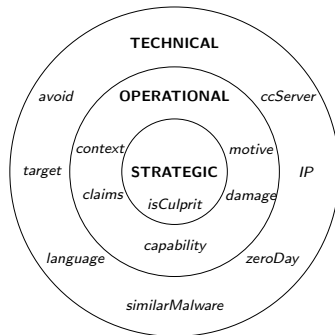
$$E = \{\text{attackSourceIP}(ip00, A_1), \text{ipGeoloc}(\text{countryC}, ip00), \text{spoofedIP}(ip00)\}$$

then the conclusion is

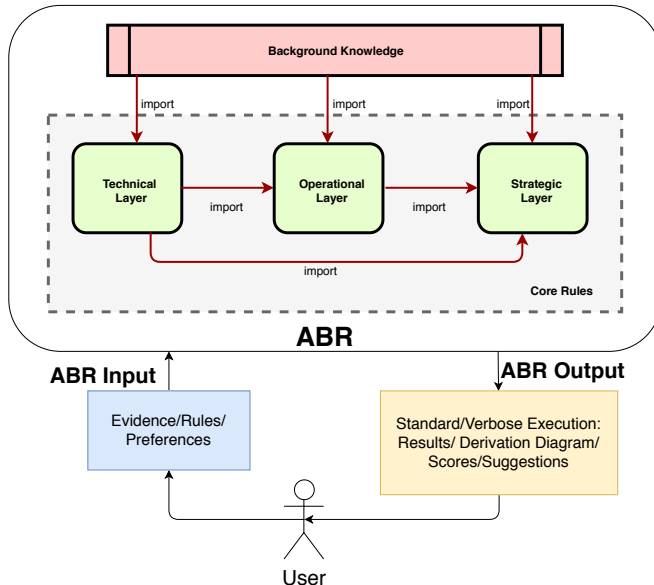
$$\neg \text{attackOrig}(\text{countryC}, A_1).$$

# Social Model used by ABR

- ABR is based on the **Q-Model**
- The Q-Model represents how the **analysts** perform the **attribution process** of cyber-attacks
- The pieces of evidence and the reasoning rules are **divided** in three **layers**



# Argumentation-Based Reasoner for Attribution



## Conclusions and Future Work

# Conclusions

- A technique to **help** the forensic investigator to analyze the cyber forensics evidence left after an attack.
- The **automatic reasoner**, which is based on abductive and argumentation reasoning, given the pieces of evidence:
  - Analyzes the evidence and derives **new information**
  - Provides **explainable conclusions** to who might be the culprit of an attack

# Future Work

- Fully automate the evidence **collection/extraction**
- **Enrich** *ABR* with reasoning rules and background knowledge
- Work with **probabilities** for the evidence and reasoning rules
- Empirical studies on the tool usability

# Questions?



e.karafili@soton.ac.uk  
sites.google.com/view/af-cyber  
cyber.southampton.ac.uk