Helping Forensics Analysts to Understand and Attribute Cyber-Attacks

Erisa Karafili

Acknowledgments

This work was supported by the European Union's H2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 746667.

Imperial College London

The future is Interconnected ...

... and comes with its own security challenges.

Missing Explanations

- ▷ Risks and Threats
- ▷ Evaluation
- ▷ Countermeasures
- ▷ Effects

Imperial College London

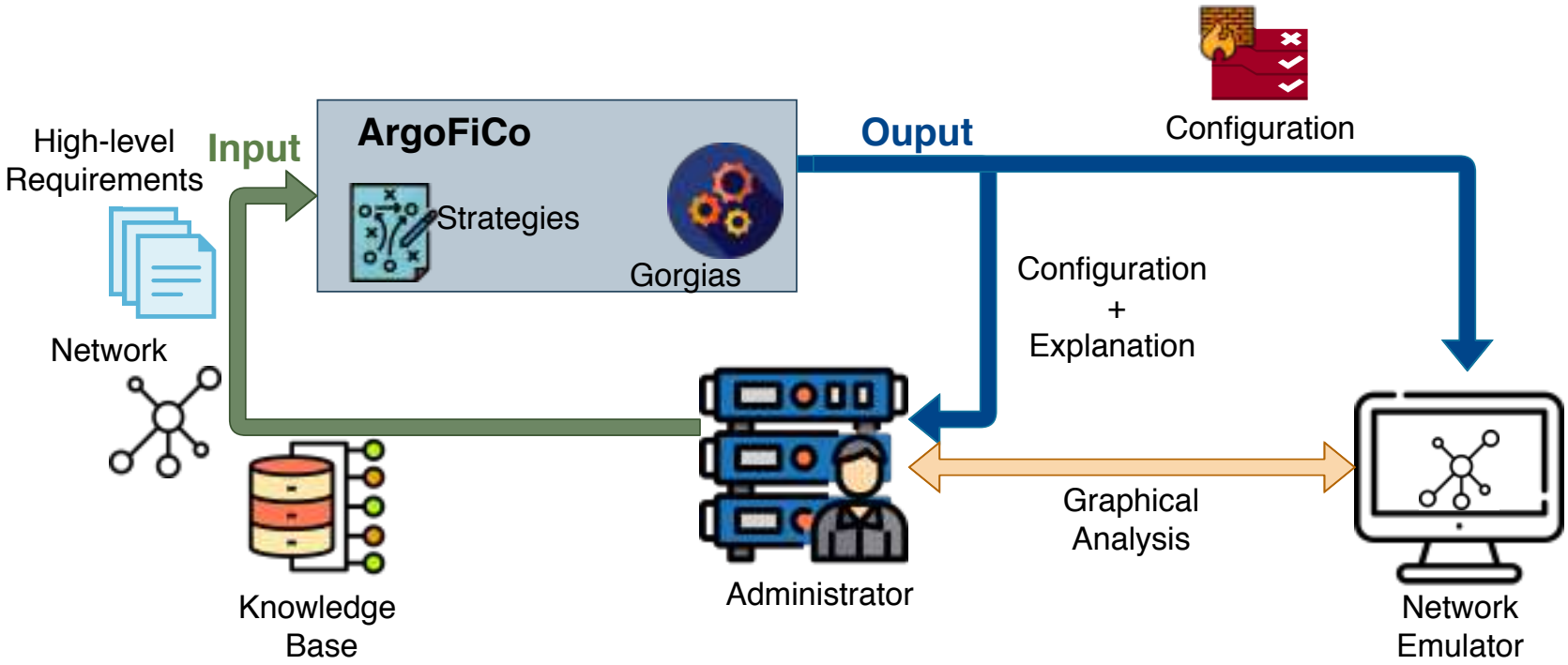
Current Solution

- ▷ Secure the systems (e.g., firewalls)
- ▷ Data centric solutions (e.g., data sharing agreements)
- ▷ Analyse Risks/Threats
- ▷ Attribution & Forensics of Cyber-Attacks

  - Not enough
  - No explanation

Imperial College London

Explaining the decision taken for the Firewall Configuration



⋆ **Karafili et al.** " *A framework for automatic firewalls configuration via argumentation reasoning*" 2019.

Imperial College London

Some Security Solutions with Explanations

Explaining Data Sharing Agreement

- ⋆ **Karafili**, Lupu "*Enabling Data Sharing in Contextual Environments: Policy Representation and Analysis*" (SACMAT 2017).
- ⋆ **Karafili et al.** "*An argumentation reasoning approach for data processing*" (Computers in Industry 2018).

Imperial College London

Attributing Cyber-Attacks

Imperial College London

Motivations

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

Imperial College London

Further Motivations

- Currently attribution is mainly performed by humans
- It suffers from human errors and is easily biased
- It is a difficult process

There is a need to:

- Help the analyst to deal with the evidence
- Explain to the analyst the provided conclusion
- Explain the used reasoning
- Provide new investigation paths

Imperial College London

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.

⋆ **Karafili**, Wang, Lupu " *An Argumentation-Based Reasoner to Assist Digital Investigation and Attribution of Cyber-Attacks*" in DFRWS EU 2020.

⋆ **Karafili**, Wang, Kakas, Lupu " *Helping Forensic Analysts to Attribute Cyber-Attacks: An Argumentation-Based Reasoner*" in *PRIMA 2018*.

⋆ **Karafili**, Kakas, Spanudakis, Lupu "*Argumentation-based Security for Social Good*" (AAAI Fall Symposium 2017)

Imperial College London

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:

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

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.

Imperial College London

A Simple Example

Given the argument pair (T, P) :

$$T = \{r_1 : attackOrig(X, Attack) \leftarrow ipGeoloc(X, IP), attackSourceIP(IP, Attack).$$
$$r_2 : \neg attackOrig(X, Attack) \leftarrow ipGeoloc(X, IP), attackSourceIP(IP, Attack), spoofedIP(IP).\}$$
$$P = \{p_1 : r_2 > r_1\}$$

and the following evidence:

$$E = \{attackSourceIP(ip1, attack1), ipGeoloc(countryC, ip1)\}$$

the conclusion is:

$$attackOrig(countryC, attack1).$$

If the evidence is:

$$E = \{attackSourceIP(ip2, attack2), ipGeoloc(countryC, ip2), spoofedIP(ip2)\}$$

then the conclusion is

$$\neg attackOrig(countryC, attack2).$$

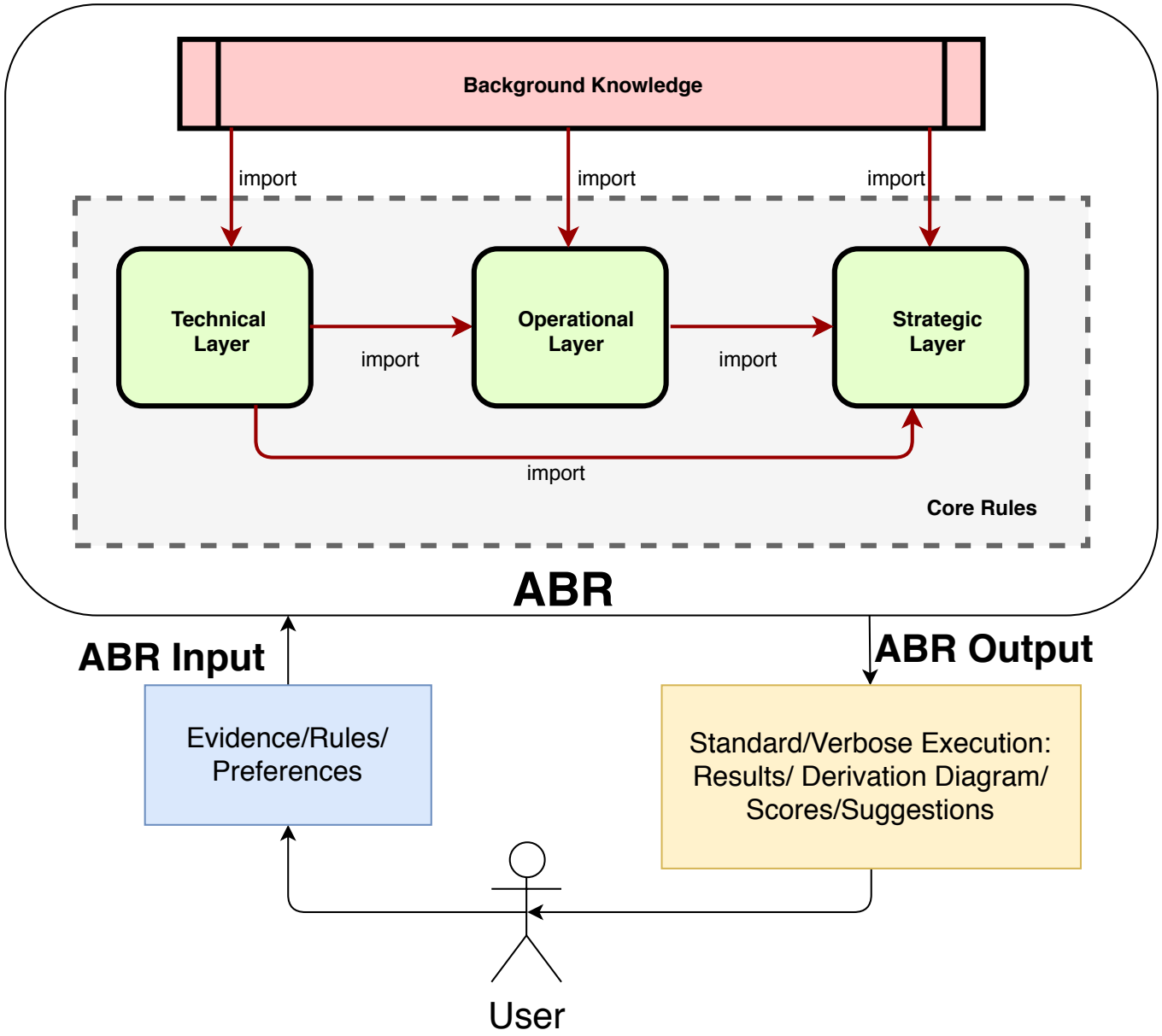
Imperial College London

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

Imperial College London

Argumentation-Based Reasoner for Attribution



Imperial College London

ABR Execution Example (1 of 4)

US bank hack occurred in 2012, where US banks faced denial of service (Dos) attacks.

- The banks' web hosting services were infected by a malware called *Itsoknoproblembro*
- *Itsoknoproblembro* hijacked the corporate clouds
- US placed economic sanctions against Iran in February 2012

$$
\begin{aligned}
&\mathit{targetCountry}(\mathit{usa}, \mathit{usbankhack}).\\
&\mathit{attackPeriod}(\mathit{usbankhack}, [2012, 9]).\\
&\mathit{malwareUsed}(\mathit{itsoknoproblembro}, \mathit{usbankhack}).\\
&\mathit{hijackCorporateClouds}(\mathit{usbankhack}).\\
&\mathit{imposedSanctions}(\mathit{usa}, \mathit{iran}, [2012, 2]).
\end{aligned}
$$

Imperial College London

Example (2 of 4)

We can assume the following:

$$
\begin{aligned}
&\mathit{targetCountry}(\mathit{usa}, \mathit{usbankhack}).\\
&\mathit{attackPeriod}(\mathit{usbankhack}, [2012, 9]).\\
&\mathit{context}(\mathit{political}, \mathit{usbankhack}).\\
&\mathit{specificTarget}(\mathit{usbankhack}).\\
&\mathit{malwareUsed}(\mathit{itsoknoproblembro}, \mathit{usbankhack}).\\
&\mathit{hijackCorporateClouds}(\mathit{usbankhack}).\\
&\mathit{imposedSanctions}(\mathit{usa}, \mathit{iran}, [2012, 2]).
\end{aligned}
$$

Imperial College London

Example (2 of 4)

$t\_1 : highLevelSkill(Att) \leftarrow hijackCorporateClouds(Att).$

$targetCountry(usa, usbankhack).$
$attackPeriod(usbankhack, [2012, 9]).$
$context(political, usbankhack).$
$specificTarget(usbankhack).$
$malwareUsed(itsoknoproblembro, usbankhack).$
$hijackCorporateClouds(usbankhack).$
$imposedSanctions(usa, iran, [2012, 2]).$

Imperial College London

Example (2 of 4)

$t\_1 : highLevelSkill(usbankhack) \leftarrow hijackCorporateClouds(usbankhack).$

$targetCountry(usa, usbankhack).$
$attackPeriod(usbankhack, [2012, 9]).$
$context(political, usbankhack).$
$specificTarget(usbankhack).$
$malwareUsed(itsoknoproblembro, usbankhack).$
$hijackCorporateClouds(usbankhack).$
$imposedSanctions(usa, iran, [2012, 2]).$
$highLevelSkill(usbankhack).$

Imperial College London

Example (2 of 4)

$t\_2 : reqHighRes(Att) \leftarrow highLevelSkill(Att).$

$targetCountry(usa, usbankhack).$
$attackPeriod(usbankhack, [2012, 9]).$
$context(political, usbankhack).$
$specificTarget(usbankhack).$
$malwareUsed(itsoknoproblembro, usbankhack).$
$hijackCorporateClouds(usbankhack).$
$imposedSanctions(usa, iran, [2012, 2]).$
$highLevelSkill(usbankhack).$

Imperial College London

Example (2 of 4)

$t\_2 : reqHighRes(usbankhack) \leftarrow highLevelSkill(usbankhack).$

$targetCountry(usa, usbankhack).$
$attackPeriod(usbankhack, [2012, 9]).$
$context(political, usbankhack).$
$specificTarget(usbankhack).$
$malwareUsed(itsoknoproblembro, usbankhack).$
$hijackCorporateClouds(usbankhack).$
$imposedSanctions(usa, iran, [2012, 2]).$
$highLevelSkill(usbankhack).$
$reqHighRes(usbankhack).$

Imperial College London

Example (3 of 4)

From the background knowledge, we have that:

$$cybersuperpower(iran).$$

$$t\_3 : hasResources(X) \leftarrow cybersuperpower(X).$$

$$t\_3 : hasResources(iran) \leftarrow cybersuperpower(iran).$$

Imperial College London

Example (3 of 4)

$targetCountry(usa, usbankhack).$
$attackPeriod(usbankhack, [2012, 9]).$
$context(political, usbankhack).$
$specificTarget(usbankhack).$
$malwareUsed(itsoknoproblembro, usbankhack).$
$hijackCorporateClouds(usbankhack).$
$imposedSanctions(usa, iran, [2012, 2]).$
$highLevelSkill(usbankhack).$
$reqHighRes(usbankhack).$
$hasResources(iran).$

Imperial College London

Example (3 of 4)

$$op\_1 : hasCapability(X, Att) \leftarrow reqHighRes(Att), hasResources(X).$$

$targetCountry(usa, usbankhack).$
$attackPeriod(usbankhack, [2012, 9]).$
$context(political, usbankhack).$
$specificTarget(usbankhack).$
$malwareUsed(itsoknoproblembro, usbankhack).$
$hijackCorporateClouds(usbankhack).$
$imposedSanctions(usa, iran, [2012, 2]).$
$highLevelSkill(usbankhack).$
$reqHighRes(usbankhack).$
$hasResources(iran).$

Imperial College London

Example (3 of 4)

$op\_1 : hasCapability(iran, usbankhack) \leftarrow reqHighRes(usbankhack),$
$hasResources(iran).$

$targetCountry(usa, usbankhack).$
$attackPeriod(usbankhack, [2012, 9]).$
$context(political, usbankhack).$
$specificTarget(usbankhack).$
$malwareUsed(itsoknoproblembro, usbankhack).$
$hijackCorporateClouds(usbankhack).$
$imposedSanctions(usa, iran, [2012, 2]).$
$highLevelSkill(usbankhack).$
$hasResources(iran).$
$hasCapability(iran, usbankhack).$

Imperial College London

Example (3 of 4)

$op\_2 : hasPolMotive(C, T, Date) \leftarrow imposedSanctions(T, C, Date).$

$targetCountry(usa, usbankhack).$
$attackPeriod(usbankhack, [2012, 9]).$
$context(political, usbankhack).$
$specificTarget(usbankhack).$
$malwareUsed(itsoknoproblembro, usbankhack).$
$hijackCorporateClouds(usbankhack).$
$imposedSanctions(usa, iran, [2012, 2]).$
$highLevelSkill(usbankhack).$
$hasResources(iran).$
$hasCapability(iran, usbankhack).$

Imperial College London

Example (3 of 4)

$op\_2 : hasPolMotive(iran, usa, [2012, 2]) \leftarrow imposedSanctions(usa, iran, [2012, 2]).$

$targetCountry(usa, usbankhack).$
$attackPeriod(usbankhack, [2012, 9]).$
$context(political, usbankhack).$
$specificTarget(usbankhack).$
$malwareUsed(itsoknoproblembro, usbankhack).$
$hijackCorporateClouds(usbankhack).$
$imposedSanctions(usa, iran, [2012, 2]).$
$highLevelSkill(usbankhack).$
$hasResources(iran).$
$hasCapability(iran, usbankhack).$
$hasPolMotive(iran, usa, [2012, 2]).$

Imperial College London

Example (4 of 4)

$$op\_3 : hasMotive(C, Att) \leftarrow \begin{array}{l} targetCountry(T, Att), \\ attackPeriod(Att, Date1), \\ hasPolMotive(C, T, Date2), \\ dateApplicable(Date1, Date2), \\ context(political, Att), \\ specificTarget(T). \end{array}$$

Imperial College London

Example (4 of 4)

$op\_3 : hasMotive(iran, usbankhack) \leftarrow$ $targetCountry(usa, usbankhack)$,
$attackPeriod(usbankhack, [2012, 9])$,
$hasPolMotive(iran, usa, [2012, 2])$,
$dateApplicable([2012, 9], [2012, 2])$,
$context(political, usbankhack)$,
$specificTarget(usbankhack)$.

Imperial College London

Example (4 of 4)

$targetCountry(usa, usbankhack).$
$\ldots$
$highLevelSkill(usbankhack).$
$reqHighRes(usbankhack).$
$hasResources(iran).$
$hasPolMotive(iran, usa, [2012, 2]).$
$hasCapability(iran, usbankhack).$
$hasMotive(iran, usbankhack).$

Imperial College London

Example (4 of 4)

$str\_1: \ isCulprit(X, Att) \leftarrow hasMotive(X, Att), hasCapability(X, Att).$

$targetCountry(usa, usbankhack).$
$\ldots$
$highLevelSkill(usbankhack).$
$reqHighRes(usbankhack).$
$hasResources(iran).$
$hasPolMotive(iran, usa, [2012, 2]).$
$hasCapability(iran, usbankhack).$
$hasMotive(iran, usbankhack).$

Imperial College London

Example (4 of 4)

$str\_1 : isCulprit(iran, usbankhack) \leftarrow hasMotive(iran, usbankhack),$
$hasCapability(iran, usbankhack).$

$targetCountry(usa, usbankhack).$
$\ldots$
$highLevelSkill(usbankhack).$
$reqHighRes(usbankhack).$
$hasResources(iran).$
$hasPolMotive(iran, usa, [2012, 2]).$
$hasCapability(iran, usbankhack).$
$hasMotive(iran, usbankhack).$
$isCulprit(iran, usbankhack).$

Imperial College London

Helping the Analyst

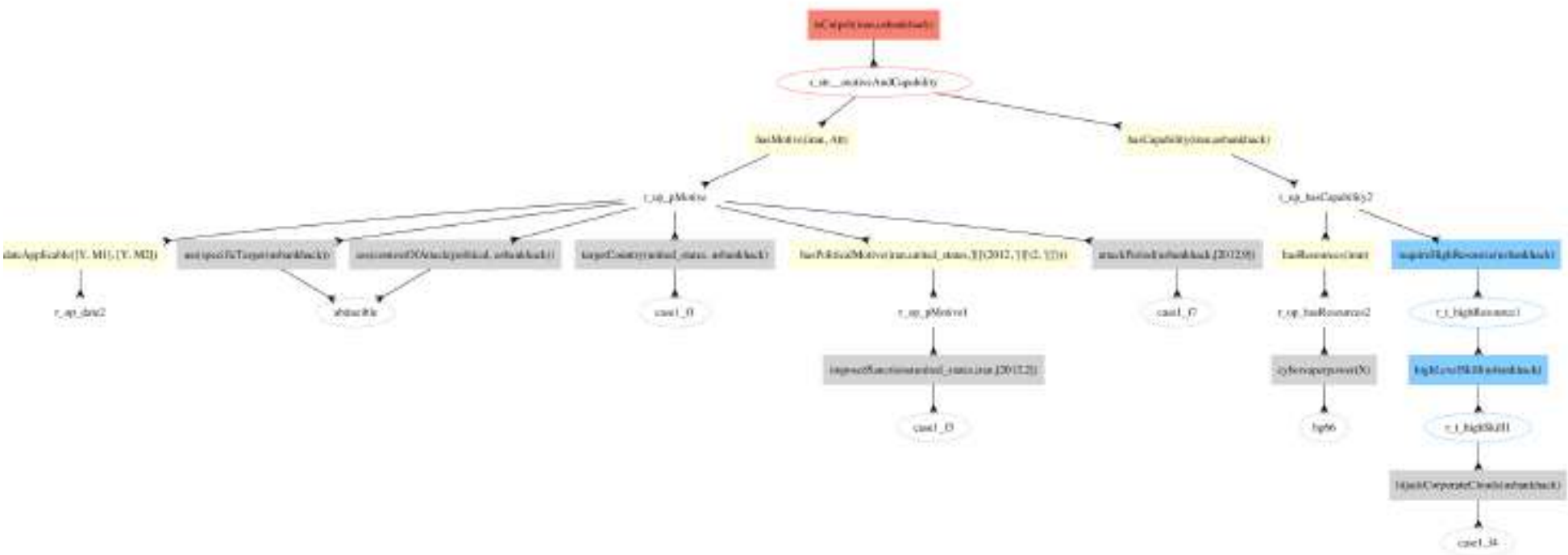
The output of ABR is the answer of the query together with:

- The used evidence and applied rules
- The abduced evidence
- Suggestions of rules and possible evidence that can prove the abduced one
- Graphical representation of the derivation
- A numerical score for the result
- The used argumentation tree

Imperial College London

Graphical Representation

- A derivation tree, where the root is the conclusion and the leaves are the evidence
- The derivation tree is color coded



Imperial College London

Scoring System

ABR provides numerical scores together with the result.

The score is calculated by summing the used evidence's score

- Score 3 to every specific-case evidence
- Score 1 to every evidence from the background knowledge
- Score 0 to the abduced evidence

The scoring system uses the *specificity criteria*.
▷ Given two results, the one that uses more specific information has a higher score.

Imperial College London

Example of the scoring system (1/2)

$$rule_1 : isCulprit(C, A) \leftarrow hasMotive(C, A), attackOrigin(C, A), country(C).$$
$$rule_2 : isCulprit(C, A) \leftarrow attackOrigin(C, A), country(C).$$

$rule_1$ should be stronger than $rule_2$ as

- C and A satisfy that satisfy $rule_1$ are a proper subset of the ones that satisfy $rule_2$.

Imperial College London

Example of the scoring system (2/2)

$$rule_1 : isCulprit(C, A) \leftarrow hasMotive(C, A), attackOrigin(C, A), country(C).$$
$$rule_3 : isCulprit(C, A) \leftarrow existingGroupClaimedResponsibility(C, A).$$

- Difficult to spot any relation between the results of these rules.
- Intuitively, $rule_1$ should be stronger than $rule_3$ as
  - $rule_3$'s body predicate is a single piece of evidence while $rule_1$ uses three pieces of evidence.
- It depends by the case study.

Imperial College London

Scoring Calculation for our Case Study

The score for our case study is 13

3	$targetCountry(usa, usbankhack)$.
0	$context(political, usbankhack)$.
0	$specificTarget(usbankhack)$.
3	$attackPeriod(usbankhack, [2012, 9])$.
3	$imposedSanctions(usa, iran, [2012, 2])$.
–	$hasPolMotive(iran, usa, [2012, 2])$.
–	$hasMotive(iran, usbankhack)$.
3	$hijackCorporateClouds(usbankhack)$.
–	$highLevelSkill(usbankhack)$.
–	$reqHighRes(usbankhack)$.
1	$hasResources(iran)$.
–	$hasCapability(iran, usbankhack)$.

Imperial College London

Different Scores for the same Result for the Case Study

- ABR provides all possible conclusions.
- The same conclusion can have different scores, (19)

$$isCulprit(iran, usbankhack).$$

- ABR uses other evidence:
  $malwareUsed(itsoknoproblembro, usbankhack)$.

Imperial College London

Argumentation Tree

An argumentation tree is provided for each result.

Imperial College London

Further Investigation

- ABR provides suggestions of other paths of investigation
- This feature permits to
  - Avoid human bias
  - Share lesson learned

Imperial College London

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 pieces of evidence
  - Provides explainable conclusions to who might be the culprit of an attack

Imperial College London

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

Imperial College London

Questions?

e.karafili@imperial.ac.uk
http://www.imperial.ac.uk/people/e.karafili
http://rissgroup.org/

RISS

Imperial College London

References

1. **Karafili**, Wang, Lupu " *An Argumentation-Based Reasoner to Assist Digital Investigation and Attribution of Cyber-Attacks*" in *DFRWS EU 2020*.
2. **Karafili**, Valenza, Chen, Lupu " *A framework for automatic firewalls configuration via argumentation reasoning*" 2019.
3. **Karafili**, Wang, Kakas, Lupu " *Helping Forensic Analysts to Attribute Cyber-Attacks: An Argumentation-Based Reasoner*" in *PRIMA 2018*.
4. **Karafili**, Spanaki, Lupu "*An argumentation reasoning approach for data processing*" in *Computers in Industry 2018*.
5. **Karafili**, Lupu "*Enabling Data Sharing in Contextual Environments: Policy Representation and Analysis*" in *SACMAT 2017*.
6. **Karafili**, Kakas, Spanudakis, Lupu "*Argumentation-based Security for Social Good*" in *AAAI Fall Symposium 2017*.

Imperial College London