A Complete Formalized Knowledge Representation Model for Advanced Digital Forensics Timeline Analysis

Yoan Chabot\textsuperscript{a,b}, Aurélie Bertaux\textsuperscript{a}, Christophe Nicolle\textsuperscript{a} and M-Tahar Kechadi\textsuperscript{b}

yoan.chabot@checksem.fr

\textsuperscript{a} CheckSem Team, Laboratoire Le2i, Université de Bourgogne, Dijon, FRANCE
\textsuperscript{b} School of Computer Science & Informatics, University College Dublin, IRELAND
Outline

Background

Problem Statement

The Underlying Issues

Existing solutions

Our solution

Formal Knowledge Model

Extraction

Analysis

Conclusion

Contributions

Future works

Outline

Problem Statement

The Underlying Issues

Existing solutions

Formal Knowledge Model

Extraction

Analysis

Contributions

Future works
**Event Reconstruction**

**Goal:** Determine the circumstances of the incident
The Underlying Issues

Technical gaps

- Large amount of data
- Heterogeneity (Semantic, Format, Time)

Digital Crime Scene

Legal requirements

- Credibility
- Veracity
- Precision
- Reproducibility
Existing Solutions

ECF, FORE, Finite state machine approach, Zeitline, Neural networks approach, CyberForensic TimeLab, etc.

*log2timeline* by Kristinn Gudjonsson

Super-timelines using a large number of sources

- Windows Event Logs
- Web Browsers Histories
- Apache logs
- PDF document metadata
- Firewall logs
- etc.

```
Yoan@Checksem-PC /cygdrive/j/Local Workspace/plaso
$ ./log2timeline ../output.dump ../Scenarios/scenario1/EnCase/scenario1.E01 > log.txt

Yoan@Checksem-PC /cygdrive/j/Local Workspace/plaso
$ ./psort -w ../timeline.txt ../output.dump > log.txt
```
Knowledge Model

How to analyze this large amount of data?
Introduce a semantically rich knowledge representation of events to enhance analysis capabilities
**Knowledge Model**

Entities

- \( s \in S = \{ a \in A_s \mid s \alpha_s a \} \)
- \( o \in O = \{ a \in A_o \mid x \alpha_o a \} \)
- \( O \subseteq \wp(A_o) \)
- \( f \in F = \{ f \in A_f \mid x \alpha_f a \} \)

- \( e \in E = \{ t_{\text{start}}, t_{\text{end}}, l, S_e, O_e, E_e \} \)
- \( S_e = \{ s \in S \mid e \in E, s \sigma_s e \} \)
- \( O_e = \{ o \in O \mid e \in E, e \sigma_o o \} \)
- \( E_e = \{ x \in E \mid e \in E, e \sigma_e x \} \)
Knowledge Model

Relationships

- $\sigma_s = \{\text{participation, repercussion}\}$
- $\sigma_o = \{\text{creation, suppression, modification, utilization}\}$
- $\sigma_e = \{\text{correlation}\}$
- $\sigma_f = \{\text{support}\}$
- $\text{support}(en \in \{E \times O \times S\}) = \{f \in F \mid f \sigma_f en\}$
Footprints left on the crime scene

Extraction and Mapping Operators

Knowledge Base

Inference Operators

Analysis Operators

Conclusions

Enhanced Timeline

Footprints

left on the

crime scene

Operators
From Raw Data To Giant Graphs

Mapping: 1st Example
Mapping: 1st Example
Mapping: 1st Example
Mapping: 2nd Example
2014-07-03T07:36:46.662000+00:00, Content Deletion Time, RECBIN, Recycle Bin, C:/Users/Yoan/Pictures/dfrws12-039.jpg, recycle_bin, TSK:/$Recycle.Bin/S-1-5-21-3724914695-4089496160-3424763353-1000/$IAXNK4E.jpg, -3, 378521

Mapping: 2nd Example
Graph Connections

Subject

Event

Object
**Automatic Analysis Operators**

**Correlation**

\[ \text{Correlation}_O(e, x) = \frac{|O_e \cap O_x|}{\max(|O_e|, |O_x|)} \]

**Object Correlation**

**Subject Correlation**

**Event Correlation**

- Temporal Correlation
- Knowledge-Based Correlation
Automatic Analysis Operators

\[ \text{Correlation}_S(e, x) = \frac{|S_e \cap S_x|}{\max(|S_e|, |S_x|)} \]
**Automatic Analysis Operators**

**Correlation**

\[ \text{Correlation}_T(e, x) = \alpha \times \text{starts}(e, x) + \alpha \times \text{equals}(e, x) + \text{meets}(e, x) + \text{overlaps}(e, x) + \text{during}(e, x) + \text{finishes}(e, x) + \text{before}(e, x) \]
Automatic Analysis Operators

Correlation_{KBR}(e, x) = \sum_{r=1}^{n} rule_{r}(e, x)

With rule_{r}(e, x) = 1 if the rule is satisfied and 0 otherwise

<table>
<thead>
<tr>
<th>Object Correlation</th>
<th>Subject Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporal Correlation</td>
<td>Knowledge-Based Correlation</td>
</tr>
</tbody>
</table>
Correlation\((e, x)\) 

\[
= \text{Correlation}_T(e, x) + \text{Correlation}_S(e, x) + \text{Correlation}_O(e, x) + \text{Correlation}_{KBR}(e, x)
\]

**Correlation**\((e_1, e_2) \approx 1,143\)

- \(\text{Correlation}_O(e_1, e_2): \) \(o_1 \Rightarrow \frac{1}{1} = 1\)
- \(\text{Correlation}_S(e_1, e_2): \) \(\emptyset \Rightarrow 0/1 = 0\)
- \(\text{Correlation}_T(e_1, e_2): \) 2014-07-03T07:36:39 \(\rightleftharpoons\) 2014-07-03T07:36:46 \(\Rightarrow \approx 0,143\)
- \(\text{Correlation}_{KBR}(e_1, e_2): \) 0
**Automatic Analysis Operators**

**Correlation**($e_1, e_3$) ≈ 0
- $Correlation_0(e_1, e_3)$: $\emptyset \Rightarrow 0/1=0$
- $Correlation_5(e_1, e_3)$: $\emptyset \Rightarrow 0/1=0$
- $Correlation_T(e_1, e_3)$: 2014-07-03T07:36:39 $\leftrightarrow$ 2010-11-20T04:58:26 $\approx 0$
- $Correlation_{KBR}(e_1, e_3)$: 0
Data volume

Automatic Operators

Heterogeneity

Unified model of knowledge representation

Extractors dedicated to each source

Credibility

Based on a formal knowledge representation

Reproducibility

Storing information about provenance

Contributions

Heterogeneity

Scalable Technologies

Credibility

Deduced from each source
Future Works

- New Analysis operators
- New Semantic dimensions
- Mechanisms for knowledge checking and reproducibility
A Complete Formalized Knowledge Representation Model for Advanced Digital Forensics Timeline Analysis

Yoan Chabot\textsuperscript{a,b}, Aurélie Bertaux\textsuperscript{a}, Christophe Nicolle\textsuperscript{a} and M-Tahar Kechadi\textsuperscript{b}

yoan.chabot@checksem.fr

\textsuperscript{a} CheckSem Team, Laboratoire Le2i, Université de Bourgogne, Dijon, FRANCE
\textsuperscript{b} School of Computer Science & Informatics, University College Dublin, IRELAND