How to Reuse Knowledge about Forensic Investigations

Lorenzo Martignoni
Università degli Studi di Milano -- Bicocca

Danilo Bruschi - Mattia Monga
Università degli Studi di Milano
Contents

- Motivations and goals
- A systematic approach to critical thinking
- Reusing forensic knowledge
- An example
- Conclusions
Motivations

- computer forensics investigations are complex because of the nature of digital evidence (volatility and skilled interpretation)
- the investigative process, in order to be presented in court, must be sound and complete, as much as possible; often every detail counts
- there are common investigative patterns that could be exploited to ease the work of investigators
Goals

★ represent the logical process followed in the proof of a thesis: critical thinking

★ record collected information in a way that ease quality assessment

★ organize past experience to foster knowledge sharing among forensic experts

★ produce reusable forensic knowledge to be used as support during investigations
The investigative process

- preliminary analysis of the case
- formulation of hypotheses on the state of the world that caused the case
- collection of evidence on the basis of these hypotheses
- correlation of actual evidence with hypotheses

preliminary analysis

hypotheses formulation

evidence collection

correlation among evidence and hypotheses

case interpretation
The investigative process

- revision of hypotheses: *abduction*
- repetition of the process until the consistency state of the knowledge about the case is acceptable
- interpretation, *by the investigator*, of the hypotheses against the collected evidence
A Cartesian approach to manage the complexity

1. **evidence**: nothing that is not clear and evident can be accepted

2. **analysis**: a complex problem should be decomposed in easier parts

3. **synthesis**: a decomposed problem has to be recomposed, verifying every partial solution

4. **enumeration**: review the whole process to verify the soundness and completeness
**Principle of evidence (1)**

- facts, observations, real things (data) to argument in favor or against a hypothesis
- conclusions have to be drawn providing tangible data
- evidence and its relevance is *context sensitive*
Principle of analysis (2)

_complex arguments ought to be separated in small ones

_the initial hypothesis is decomposed in sub-hypotheses:

\( H \rightarrow H_1, H_2, H_3, H_4, \ldots, H_n \)

_""," is not a logical connective and "→" is not a logical equivalence_
Principle of synthesis (3)

★ recomposition of the partial solution of the decomposed problem

★ from a forensic viewpoint: “collecting information to prove or disprove the occurrence of an event in the real world”

$$H_i \Rightarrow E_1, E_2, E_3, \ldots, E_n$$

★ “⇒” denotes the application of tests in order to evaluate the hypothesis
**Principle of synthesis (3)**

- Every evidence collection test will lead, if applicable, to a success or a failure.
- The set of applicable tests is by no means complete.
- Sometimes highly relevant tests cannot be performed.
- The strength of each test and the correlation among several of them is not a constant but context sensitive.
**Principle of enumeration (4)**

- by making the process explicit is possible to assess the quality of the whole process
- reuse of past experience in analysis and synthesis decreases the possibility of human errors and omissions

Collected information can be organized as *forensic graph*
*Forensic graph*

\[
FG = \langle H, E, F_{h}, F_{e}, w \rangle
\]

A DAG where:

- \( H \) set of hypotheses
- \( E \) set of evidences
- \( F_{h} \) decomposition relation (\( F_{h} \subseteq H \times H \))
- \( F_{e} \) association relation (\( F_{e} \subseteq H \times E \times w \))
- \( w \) weight of evidence (\( w \in \{+, -, ?\} \))
Forensic graph

- used to represent all the knowledge acquired over the time
- hypotheses and evidences are represented in natural language
- expresses the relations among hypotheses and evidence relevant for their validity
- every case is instantiated in a **case graph**
Forensic graph
Case graph

- Case graph models logic behind the detective's analysis in a specific criminal case.
- A new graph is built using only the hypotheses and evidence related to the current context.
- The weight of evidence expresses how evidence affects an hypotheses (corroboration, contradiction or the test was not performed).
Learning

- during the construction of case graph new hypotheses can be formulated
- new link among hypotheses and evidence can be discovered
- forensic graph is updated to reflect the new experience
- current experience will be available for future case
Reusing past experiences and learning

New case

Old case

Solved case

Knowledge
An example

$H$: email account $\text{bob@domain}$, registered by user $\text{Bob}$, has been used to send a harmful message $M$, to user $\text{V}$. $\text{Bob}$ is the author of $M$ and its sender.
An example (hypothesis decomposition)

$H_1$: Bob has sent message $M$ from his computer $C$

$H_2$: sendmail, the mail transfer agent installed on $C$, has been configured to use $bob@domain$ as the From: header

$H_3$: when $M$ was sent ($T$), $C$ has been in use

$H_4$: when $M$ was sent ($T$), $C$ was connected to the Internet
$H_3$: when $M$ was sent ($T$), $C$ has been in use

$E_1$: are there files modified, created, deleted, accessed at time $T$?

$E_2$: are there files that contain information about user activity (browser history, email-client recent file list, ...) at time $T$?

$E_3$: are there files that contain information about system activity (events logs, applications logs, ...) at time $T$?
$H_3$: when $M$ was sent $(T)$, $C$ has been in use

$E_1$: not found

$E_2$: found

$E_3$: N/A (logs were encrypted)

Is evidence conclusive or inconclusive? The answer is left to the investigators!
Limitations

- it is neither possible to express nor to evaluate how much an evidence influences an hypothesis: *inferential drag*

- expression of hypotheses and evidence in a natural language limits automatic search inside knowledge
Conclusions

- argumentations supporting a hypothesis are open to criticism
- representation through a graph renders knowledge reusable (even of subgraphs)
- knowledge can be improved as investigation experience grows
Future works

🔹 we are implementing a tool that applies our approach to be used as a guideline both for detectives and attorneys

🔹 provide a structured language to describe evidence and hypotheses in order to process them automatically

🔹 estimate the relevance of hypotheses studying the outcome of previous and concluded case: analysis of causality
Questions and suggestions are welcome...