Online Banking Malware Ontology

R. Carvalho, J.R.C. Nurse, M. Goldsmith

University of Oxford — Cyber Security Centre

Malware Investigation

Law Enforcement Agencies must cope with evidence from multiple different sources, a shortage of skilled cybercrime agents and inefficient data exchange between stakeholders.

Among the digital offences most frequently investigated, there are phishing attacks aimed at spreading financial services-targeted malware.

Such malicious software is normally sold to, and used by, multiple thieves, which vastly outnumber the amount of developers.

Arresting malware programmers, albeit more challenging, could have a cascade effect on the Online Banking Malware (OBM) ecosystem.

Therefore, novel investigation techniques balancing cybercrime domain knowledge and computer processing power should be researched.

Applying Ontologies

Semantic technologies could enhance relationship discovery and hypothesis testing within the OBM domain.

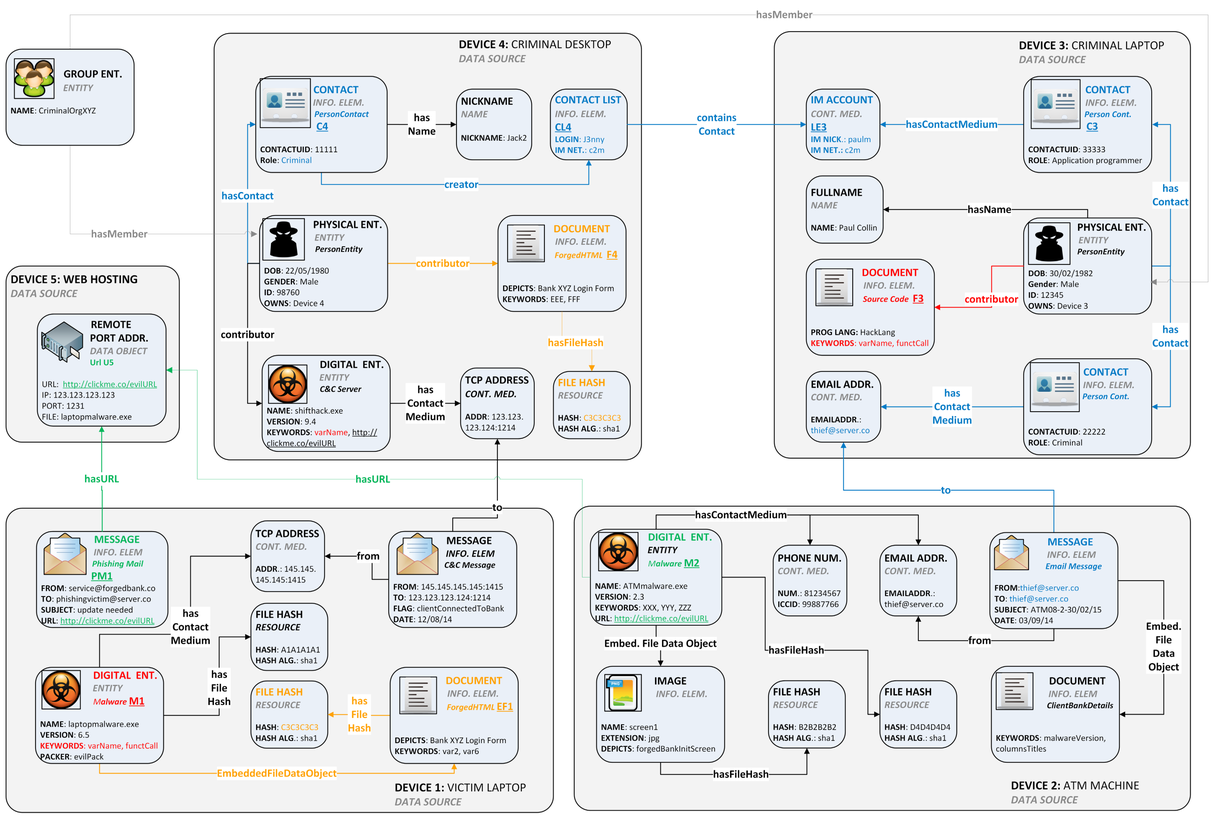
For instance, supporting analysts in reasoning among a large volume of scattered, supposedly unrelated evidence.

By defining concepts and their relationships, properties and value types, an ontology allows:

- Analysts and computer software to share a common domain understanding, enhancing interaction and information exchange;
- The reuse of domain knowledge: e.g. file hash concepts are also useful for investigating other cybercrimes;
- To easily devise domain assumptions: e.g. “malware family X has at least 3 of their embedded files identical”.

Main Objectives

- Link different criminal organizations
- Identify malware developers
- Facilitate future data integration
- Optimize cyber evidence retrieval



OBM Concepts, Relationships and Properties

Icons by Hopstarter / CC BY-NC-ND

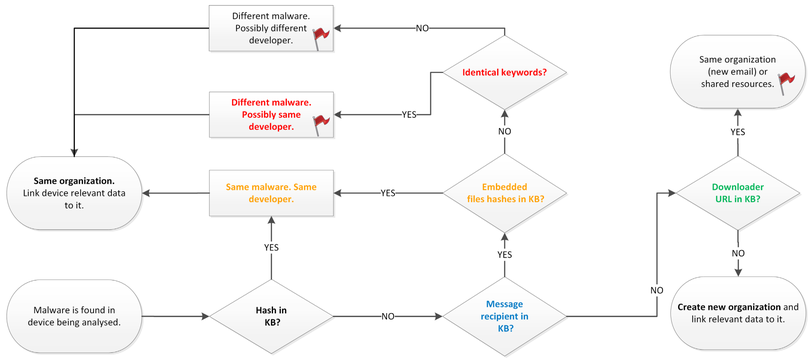
Device 1 x Device 4
AXIOMS: Malware (M) is a file (F). All files have hashes. Some files have embedded files (EF). Embedded files are files.
Malware M1 contains Embedded File EF1.
EF1 hash matches File F4 hash.
HYPOTHESIS: If one malware embedded file has been found as a file somewhere else, its contributor is at least a script kiddie.

Device 1 x Device 2
AXIOMS: Malware can be downloaded from a remote computer. Malware can connect to a external URL (U).
All embedded file hashes from Malware M1 and M2 differ.
Phishing mail PM1 and M2 contain references to URL U5.
HYPOTHESIS: Different malware connecting to the same URL belong to the same criminal organization.

Device 1 x Device 3
AXIOMS: Files have keywords. Some keywords are unique.
Malware M1 contains keywords varName, functCall.
Keywords varName, functCall were found in source code F3.
HYPOTHESIS: A non-executable file containing multiple keyword matches with a malware is its source code.

Device 3 x Device 4
AXIOMS: An entity has one or more contacts/profiles (C). Contacts might create contact lists (CL) containing entries (LE).
Contact List CL4 from Criminal Contact C4 stores List Entry LE3.
LE3 belongs to Contact C3, with role Application Programmer.
HYPOTHESIS: An application programmer who appears in the contact list of a known criminal may be a malware developer.

Information linking hypotheses



OBM Inference flowchart

Current and Future Work

A working prototype with a reduced set of the most relevant concepts relating to the malware (contact medium, URL, keyword and hash) is being implemented.

Next, concepts from entities (contact, nickname, user login and messages exchanged) will be incorporated into the implementation.

Future work will:

- Load real data into the knowledge base to validate and improve the inference rules’ efficiency and efficacy;
- Refine the approach to malware string matching by analysing previous research and sandbox platforms;
- Research Natural Language Processing techniques to automate entity extraction and provide content-based categorisation;
- Create an online resource to foster OBM ontology discussion and improvement among Law Enforcement Agencies.

References

- S. L. Garfinkel, ‘Digital forensics research: The next 10 years’, Digital Investigation, 2010;
- N . F. Noy, D. L. McGuinness, et. al.‘Ontology development 101: A guide to creating your first ontology’ 2001;
- Anti-Phishing Working Group, ‘Phishing activity trends report - 1st quarter 2014’;
- Nepomuk Consortium, ‘OSCAF ontologies’. Available: *http://www.semanticdesktop.org/ontologies/*.

Email: rodrigo.carvalho@cs.ox.ac.uk
Tel: +44 1865 610805
URL: http://www.cs.ox.ac.uk/people/rodrigo.carvalho/

The author’s DPhil programme is funded by CAPES-CSF and supported by the Brazilian Federal Police.