Environmental Scanning and Knowledge Representation for the Detection of Organised Crime Threats

BREWSTER, Benjamin <http://orcid.org/0000-0003-3536-5862>, ANDREWS, Simon <http://orcid.org/0000-0003-2094-7456>, POLOVINA, Simon <http://orcid.org/0000-0003-2961-6207>, HIRSCH, Laurence <http://orcid.org/0000-0002-3589-9816> and AKHGAR, Babak <http://orcid.org/0000-0003-3684-6481>

Published version

BREWSTER, Benjamin, ANDREWS, Simon, POLOVINA, Simon, HIRSCH, Laurence and AKHGAR, Babak (2014). Environmental Scanning and Knowledge Representation for the Detection of Organised Crime Threats. In: HERNANDEZ, Nathalie, JÄSCHKE, Robert and CROITORU, Madalina, (eds.) Graph-Based Representation and Reasoning. Lecture Notes in Computer Science (8577). Springer International Publishing, 275-280.

Copyright and re-use policy

Environmental Scanning and Knowledge Representation for the Detection of Organised Crime Threats

Ben Brewster, Simon Andrews, Simon Polovina, Laurence Hirsch, Babak Akhgar

CENTRIC, Sheffield Hallam University
{B.Brewster, S.Andrews, S.Polovina, L.Hirsch, B.Akhgar} @SHU.ac.uk

Keywords: ePOOLICE, Environmental Scanning, Open-Source Intelligence (OSINT), Formal Concept Analysis, Conceptual Graphs, Ontological Knowledge Representation.

Abstract.
ePOOLICE aims at developing an efficient and effective strategic early warning system that utilises environmental scanning for the early warning and detection of current, emergent and future organised crime threats. Central to this concept is the use of environmental scanning to detect 'weak signals' in the external environment to monitor and identify emergent and future threats prior to their materialization into tangible criminal activity. This paper gives a brief overview of the application of textual concept extraction and categorization, and the Semantic Web technologies Formal Concept Analysis and Conceptual Graphs as part of the systems technological architecture, describing their benefits in aiding effective early warning.

1 Introduction

As a result of the popularity and near ubiquitous nature of the Internet, organised crime has become ever more diverse in nature [1]. Criminal groups increasingly benefit from increased levels of collaboration, mobility around the EU and access to dynamic infrastructure, enhancing the capacity and capability of their criminal practices [2]. By their very nature, criminal groups are constantly seeking to exploit new avenues in order to sustain their illicit practices. These include, but are not limited to; drug crimes such as dealing, trafficking, cultivation and trafficking in human beings [3]. Numerous reports have studied and discussed the factors that facilitate and enable organised crime [2], [4], with others researching the mining of open-source data in order to detect communities [5] and the utility of information fusion in the detection of weak signals and the provision of strategic early warning [6].

The requirement to integrate the concepts cited has been identified in order to develop an approach to provide strategic early warning of organised crime. Although the system proposed here does not provide a comprehensive solution, it does demonstrate as a prototype how environmental scanning and semantic web technologies can be applied in order to enhance Law enforcement agencies capability in the early, strate-

gic identification combatting such threats [7], [8]. ePOOLICE proposes the development of a prototype environmental scanning system, applying a variety of state-of-the-art technological solutions, including Formal Concept Analysis and ontological knowledge representation through the use of Conceptual Graphs. At this stage it is important to note that although this paper focuses on the application of these technologies specifically, it does not form an accurate, holistic representation of the ePOOLICE project in its entirety. When considering the application of open-source scanning in this way, public privacy and surveillance fears are a key concern that must be accounted for [9]. In order to preserve the privacy of citizens, ePOOLICE refrains from the identification of specific individuals and instead focuses on the identification of patterns and observations ensuring that a 'Privacy by Design' approach to systems development is followed. For a more in depth discussion of the privacy and ethical considerations related to ePOOLICE, please see [10]. ePOOLICE integrates environmental crawling via the application of open-source scanning alongside a semantic knowledge repository, for the storage and retrieval of new and existing domain knowledge. Although not strictly part of the systems overall architecture, open-sources underpin the analytical capability of ePOOLICE through providing access to indicators, sentiment, location data and other potentially relevant concepts that may provide information that aids in increasing threat awareness. A variety of text mining and analytical approaches will be used to extract information and meaning from a number of disparate, un-structured and structured, open-source repositories. Utilising techniques that process and parse textual data in real time, the system aims to inform decision makers to assist in combatting not only current and emergent organised crime threats but also in assessing the potential for future threats.

Central to ePOOLICE, is the input from, and collaboration with end user partners from the law enforcement domain. These partners consist of law enforcement agencies themselves, criminologists and other domain experts. The knowledge provided by the end-user partners will guide the multidisciplinary, pan-European research team in the identification of poignant, current and future organised crime issues.

2 Objectives

At its core, the ePOOLICE system aims to address two main requirements;

1. The detection of organised crime: This is the early detection of existing/current and emerging organised criminal threats and criminal organisations.
2. The prediction of future organised criminal threats: Using environmental scanning it is possible to strategically assess the potential for future threats based upon historical data, patterns, and indicators derived from open-sources.

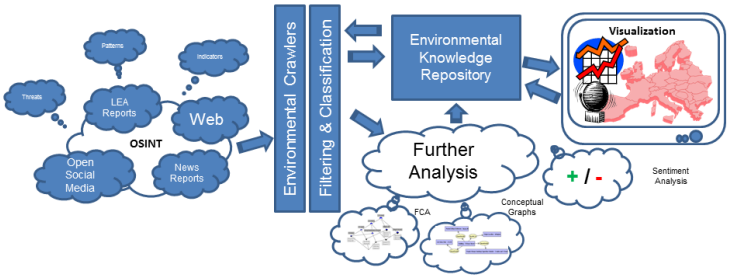
These objectives guide the principal design of ePOOLICE and are achieved through the use of what CISC (Criminal Investigation Service Canada) [11] define as 'temporal proximity'. In this definition 'primary' indicators refer to information that may be directly related to, or that occur as a result of, an organised crime (OC) threat. 'Secondary' indicators however consist of information that enables or promotes the potential for an OC threat to occur. Secondary indicators are likely to consist of in-

formation that may be identified during a PESTLE (political, economic, sociological, legal and environmental) analysis such as the identification of features including economic or political instability and legislation changes; factors that could enhance the potential for organised crime and/or increase criminal entities capacity to commit it.

3 System Architecture

Three of the key components of ePOOLICE's technical architecture (as shown in figure 1) are the environmental knowledge repository (EKR) and data fusion and analysis tools (further analysis). These components are then integrated and represented using a number of visualisation tools displaying threat indicators geographically along with subsidiary information such as trends and meta-data detailing the source and validity of indicators; information that can be applied by decision makers.

Fig. 1. ePOOLICE technical architecture overview



3.1 Data Acquisition, Content Categorization & Concept Extraction

The data acquisition aspect of ePOOLICE is concerned with the crawling and extraction of information from disparate, structured and unstructured, open sources. During extraction, data is parsed and normalised into a structured, unified format (such as XML or JSON), removing sensitive data that may contribute to victims being directly identifiable thus enabling all data to be processed and analysed via a single document pipeline, which is then marked up in preparation for further textual mining, classification and visualisation.

Post extraction, data is categorized, and key concepts identified to discern the data's relevance to specific subject areas, events and geographic locations. In addition, potential indicators or characteristics of illicit activities (both 'primary' and 'secondary" in nature) are extracted. This processing is conducted using statistical and rule based textual analysis techniques that utilize technologies such as natural language processing (NLP) to discern meaning from disparate content through the assessment of structure, linguistic patterns and concept references in the source data. For this purpose, a combination of linguistic rules, regular expressions and Boolean syntax will be applied to extract relevant concepts and identify content categories from within crawled data. For example, a newly published report such as EUROPOL's serious organised crime threat assessment [12] may contain pertinent information such as new

transit routes, transportation methods and types of exploitation that can be identified using regular expressions and Boolean syntax. Relevant indicators, terminology, known locations and routes, alongside other domain relevant knowledge is derived from the expertise of end users and semantically translated to form the taxonomy and ontology components of the system. Figure 2 gives an example of named entities that may be extracted in this way using an extract from a EUROPOL case study [13].

Fig. 2. Concept Extraction

<DATE>On 26 March 2012</DATE>, a highly organised <CRIME>drug trafficking</CRIME> network was brought to trial in <LOCATION>Sweden</LOCATION>. Eight members of the group faced criminal charges for trafficking multi-tonne shipments of high-quality <DRUG>cocaine</DRUG> from <ORIGIN>South America</ORIGIN> to <DESTINATION>Europe</DESTINATION>. Another trial on the <CRIME>money laundering</CRIME> .....

3.2 Environmental Knowledge Repository: Conceptual Graphs

Conceptual Graphs enable the representation of knowledge in a format that is discernable not only by humans but also by software and capture knowledge through the use of an ontological vocabulary [14]. In ePOOLICE, conceptual graphs are utilised to tangibly represent environmental domain knowledge using the semantic web. Formats such as RDF and OWL are used to house this knowledge within the EKR triple-store. The domain knowledge represented using conceptual graphs, corresponds directly to the taxonomy used to extract and categorize data during its acquisition. The domain knowledge embedded in the EKR provides a model upon which the system can use to effectively 'understand' crawled data sources, defining the relationships between valuable concepts such as locations and indicators.

Conceptual Graphs add value through the identification of patterns in the underlying concepts that can be identified in order to relate data to other data in the conceptual graph's vocabulary. In ePOOLICE, these value-adding features enable insights to be derived through the identification of relationships between weak indicators, which, in isolation may not seem in any way related to organised crime activity. However, by modelling existing domain knowledge using conceptual graphs, it may be possible to project and therefore discern that several weak indicators together constitute a valid indicator of illicit organised crime activity.

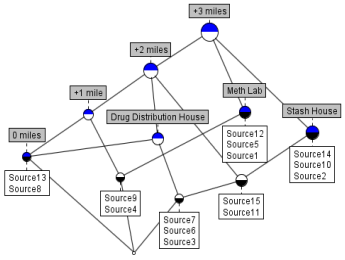
3.3 Situation Assessment: Formal Concept Analysis

One aspect of ePOOLICE's analytical armoury is Formal Concept Analysis (FCA), a semantic data analysis method that captures, categorizes and delivers data meaning in real time to influence decision makers through ontologically modelling the relationships between objects and their attributes [15]. FCA presents the sources of organised crime indicators as formal objects. These sources may be made up of data from open-sources and police reports. The indicators themselves are presented as formal attributes, with attributes forming the characteristics of objects. Attributes are made up of

named entities such as the identification of location, time and information such as drug references, thus enabling situation assessment. As a result, formal concepts represent frequent groups of indicators along with situation assessment information. The more frequent the group, the more weight can be given to the evidence. Situation assessment allows information to be appropriately visualised depending on the requirements of the analyst. For example, a map-based system to represent threat indicators geographically may form the basis of one such approach to visualisation. This can also include a measure displaying the frequency level itself, giving the analyst control over the levels of support that they are interested in.

In FCA, the scaling of continuous attributes, such as geospatial (see Figure 3) and temporal values gives the analyst control over situation assessment. Applying a 'zoom' like functionality allows the analyst to see a more aggregated, strategic perspective when 'zoomed out', while 'zooming in' enables the focus to be concentrated on the scenarios that make up events or sources, enabling the manual assessment of extreme or erroneous data. In temporal terms, 'zooming out' gives a more strategic, historical view of events, with 'zooming in' focusing solely on the current situation.

Fig. 3. FCA scaling of geospatial information



4 Concluding Remarks

ePOOLICE applies a variety of novel, state-of-the-art technologies in order to inform decision makers of current emergent, and potential future organised crime threats. Central to this capability are the semantic web technologies; FCA and Conceptual Graphs. In this overview we have presented a brief introduction to ePOOLICE, identified the rationale behind the project and described the key technological components that underpin the systems architecture.

Disclaimer

In this document, terms indicating origin or ethnicity are not being used to imply anything general or stereotypical of that origin or ethnic group. Such terms are only used as instances of factual reporting and are not be taken as a reference to any race or ethnic group as a whole. Nevertheless, for a system to monitor organized crime to operate effectively, the identification of certain elements such as gender, nationality and ethnicity, in addition to the explicit identification of crime gangs, victim groups and modus operandi are often important. For instance, Vietnamese victims are traf-

ficked into Europe (supported by several reliable sources) and, therefore, the reference to the Vietnamese origin of criminal groups is crucial to investigate such cases as it forms a key characteristic of the phenomena described. Furthermore, when sensitive or personal data is being handled, it will be done so in accordance with laws protecting the privacy and human rights of individuals, including data protection laws.

Acknowledgement
This project has received funding from European Union Seventh Framework Programme FP7/2007 - 2013 under grant agreement nº FP7-SEC-2012-312651.

5 References

1. A. Wright, *Organised crime*, Routledge, 2013.
2. Europol. *Organised Crime Threat Assessment* [Online]. available: https://www.europol.europa.eu/sites/default/files/publications/octa2011.pdf.
3. Crown Office & Procurator Fiscal Service. *Solicitor General launches Serious and Organised Crime Division* [Online]. available: http://www.copfs.gov.uk/media-site/media-releases/243-solicitor-general-launches-serious-and-organised-crime-division-socd.
4. H. Abadinsky, *Organized crime*, Cengage Learning, 2009.
5. L. Tang and H. Liu, "Community detection and mining in social media," *Synthesis Lectures on Data Mining and Knowledge Discovery*, vol. 2, no. 1, pp. 1-137 2010.
6. G.S. Ng, C. Quek and H. Jiang, "FCMAC-EWS: A bank failure early warning system based on a novel localized pattern learning and semantically associative fuzzy neural network," *Expert Syst.Appl.*, vol. 34, no. 2, pp. 989-1003 2008.
7. M. Zenko and R.R. Friedman, "UN early warning for preventing conflict," *Int.Peacekeeping*, vol. 18, no. 1, pp. 21-37 2011.
8. ePOOLICE. *ePOOLICE - About* [Online]. available: https://www.epoolice.eu/EPOOLICE/about.jsp.
9. D. Omand, J. Bartlett and C. Miller, "Introducing social media intelligence (SOCMINT)," *Intelligence and National Security*, vol. 27, no. 6, pp. 801-823 2012.
10. A. Gerdes, H.L. Larsen and J. Rouces, "Issues of Security and Informational Privacy in Relation to an Environmental Scanning System for Fighting Organized Crime," in *Flexible Query Answering Systems*, Anonymous : Springer, 2013, pp. 155-163.
11. Criminal Intelligence Service Canada (CISC). *Strategic Early Warning for Criminal Intelligence* [Online]. available: http://www.cisc.gc.ca/products_services/sentinel/document/early_warning_methodology_e.pdf.
12. EUROPOL. *EU Serious and Organised Crime Threat Assessment* [Online]. available: https://www.europol.europa.eu/sites/default/files/publications/socta2013.pdf.
13. EUROPOL. *EU Drug Markets Report: A Strategic Analysis* [Online]. available: https://www.europol.europa.eu/content/eu-drug-markets-report-strategic-analysis.
14. J.F. Sowa, *Conceptual structures: information processing in mind and machine*, Addison-Wesley Longman Publishing Co., Inc., 1984.
15. R. Wille, "Formal Concept Analysis as Mathematical Theory of Concepts and Concept Hierarchies," in B. Ganter, G. Stumme and R. Wille, Eds: Springer Berlin Heidelberg, 2005, pp. 1-33.