Towards Fully Integrated Real-time Detection Framework for Online Contents Analysis - RED-Alert Approach

Syed Naqvi, Ian Williams, Sean Enderby
School of Computing and Digital Technology
Birmingham City University
Birmingham, United Kingdom
(Syed.Naqvi, Ian.Williams, Sean.Enderby)@bcu.ac.uk

Peter Pollner, Daniel Abel
Statistical and Biological Physics Research
Hungarian Academy of Sciences
Budapest, Hungary
(pollner, abel)@elte.hu

Berta Biescas
INSIKT Intelligence
Barcelona, Spain
berta@insiktintelligence.com

Oscar Garcia
Information Catalyst for Enterprise
Valencia, Spain
oscar.garcia@informationcatalyst.com

Monica Florea, Cristi Potlog
SIVECO
Bucharest, Romania
(Monica.Florea, Cristi.Potlog)@siveco.ro

***Abstract*—Social media is extensively used nowadays and is gaining popularity among the users with the increasing growth in the network capacity, connectivity, and speed. Moreover, affordable prices of data plans, especially mobile data packages, have considerably increased the use of multimedia by different users. This includes terrorists who use social media platforms to promote their ideology and intimidate their adversaries. It is therefore very important to develop automated solutions to semantically analyse online contents to assist law enforcement agencies in the preventive policing of online activities. A major challenge for the social media forensic analysis is to preserve the privacy of citizens who use online social networking platforms. This paper presents results of European H2020 project RED-Alert that aims to enable secure and privacy preserving data processing; hence the malicious content and the corresponding personality can be ethically tracked. We have mined seven social media channels for content and providing support for ten languages for analysis. Our proposed solution is designed to ensure security and policing of online contents by detecting terrorist material. We have used social network analysis, speech recognition, face and object detection besides audio event detection to extract information from online sources that are fed in a complex event processor. We have discussed the challenges and prospects of this work especially the need of analysing online contents while respecting European and national data protection laws notably GDPR.**

***Index Terms*—Contents analysis, social network analysis, multimedia forensics, complex event processing, data protection.**

I. Introduction

Analysis of online contents analysis is very challenging from technological and ethical point of views as well as due to the scale and scope of the social media [1]. Privacy and data protection requirements call for careful consideration of the techniques and algorithms for data extraction and its processing by the analysts. Efficient solutions are needed to ensure that information is timely found with minimal chances of errors (false positives or negatives). Predicting an imminent terrorist activity without any prejudice in almost real-time is the focal point of modern policing [2] also called Policing 4.0 [3].

We have outlined the requirements of online contents analysis in the Section 2. We are working on these requirements under the auspices of European project RED-Alert that deals with the challenges of preventive policing of online activities. An overview of this project is given in the Section 3 of this paper. We have developed a Semantic Multimedia Analysis (SMA) Tool as part of the RED-Alert project. This tool is designed to extract data from multimedia contents. Its details are provided in the Section 4. Our Social Network Analysis (SNA) module is presented in the Section 5. The overall integration of these modules in the RED-alert architecture is made through a Complex Event Processing (CEP) unit. The RED-Alert CEP is described in the Section 6.

We have made a pragmatic discussion in the Section 7 about the features and issues with the RED-Alert solution and how we have managed them in the RED-Alert project. Finally we have drawn some conclusions together with a concise description of our future directions in the Section 8.

II. Requirements of Online Contents Analysis

Automatic Natural Language Processing (NLP) methodologies applied to online text play a major role in the Red Alert solution. One of the features that contributes to the triggering of alerts in the Red Alert engine bases on the automatic detection of text, whose content has high probabilities of being terrorism-related. For this purpose, three features are extracted from the text that inputs the system: i) key-ideas extraction, ii)

The research leading to the results presented in this paper has received funding from European Commission Horizon 2020 (H2020) Programme under Research and Innovation Action H2020-SEC-12-FCT-2016 (Technologies for prevention, investigation, and mitigation in the context of fight against crime and terrorism). Grant agreement number 740688.

topic extraction, and iii) terrorism-related probability calculation. The key ideas feature is analysed following a standard approach, while the topic classification and the terrorism-related probability are based on deep learning methodologies and innovative multi-language alignment applications. All the coding is performed with open-source python libraries: Scikit, Gensim, Nltk, Muse and Vecmap. NLP analysis is based on a compilation of online data, which follows GDPR requirements. The text acquired from social networks is downloaded by automatic scripts based on search terms selected by LEAs. The downloaded fields are date and text (no username, no id, no location) to guarantee complete privacy. In this first step, there is no need for any human supervision. The data is filtered using a k-anonymization. In the last step, data is saved on an encrypted server. The description of the methods applied for the three feature extractions follows.

Key ideas method extracts from the text particular patterns of words. The method works by performing POS tagging on the text and then extracting the words that match with the patterns. The specific patterns depend on the language, but are combinations such as noun plus adjective. This method receives a clean tokenized text, including stopwords, and returns a list of key ideas.

There are several existing word embedding models that are shared on the internet and that can be assimilated by scikit functions. These models have the advantage that they are already available and the disadvantage that they are not specific for terrorism-content domain, which means that the vector space does not sufficiently cover the specific lexicon that appears in the texts that RED-Alert system is interested in. Previous studies have proved that online language often significantly differs from the formal one in, for example, the used of multi-language expressions, the used of misspelling words and specfic terms used to distinguish or publicize specific subgroups or communities of users (references?). Therefore, Red Alert embedding models are trained using online terrorist journals and texts used to spread this specific ideology. Comparisons between generic word embedding models and the specific domain ones, created by RED-Alert partners, show that results are always better using the specific ones. Therefore, in languages with enough available online text, i.e. English, specific-domain word embedding models have been created and in the rest of the languages, the generic ones are used with an alignment methodology that is described in this section. Word2Vec algorithm is used for the creation of the word embedding models. This algorithm is integrated in a python script with the Gensim library. The final selected parameters applied in the embedding training are: number of occurrences = as many as frequent terms are found in the corpus, size of the layers = 300 and the Skip-Gram training algorithm.

The creation of specific domain word embedding is very good when a relatively large amount of text is available. This is the case, for example, for English. However, for other languages, data acquisition is rather more difficult. To address this issue an unsupervised method for extending the classifier to new languages is developed, transferring the available knowledge from data-rich languages to data-poor languages. Concretely, a standard data-poor embedding is rotated to the same vector space to a data-rich embedding and the supervised learning of the data-rich language is used to calculate the terrorism-related probability of the data-poor language. The alignment is done using a small dictionary between both languages.

The topic classification method is achieved in the deep learning domain. Texts are mapped to numerical vectors and are classified into the closest topic. The distances between the vectorised text and the topics are calculated with the cosine distance algorithm. Topics are selected by the users based on related terms. Results are control by a threshold level, if the calculated distance is shorter than the threshold, the closest topic is output, and in case the distance is longer than the threshold, no topic is returned.

The terrorism-related probability is calculated with an ExtraTree algorithm designed with supevised learning. A data set resulting from a compilation of online texts with a binary human annotation - classes: terrorism-related and not terrorism related - is used for the supervised learning.

III. RED-ALERT PROJECT

This work is a part of European Project **RED-Alert** (**R**eal-time **E**arly **D**etection and **Alert** System for Online Terrorist Content) [4]. RED-Alert solution aims to be a decision maker system for helping Law Enforcement Agencies (LEAs) to control the online terrorist activity, saving time and human resources and increasing the effectiveness. The solution provides a complete vision of online activities by integrating different methods, which analyse images, text and social networks on real time. The system triggers an alert based on the combination of weight probabilities coming from the different analysis (Figure 1).

This project is consist of 16 partners from European Unions Member States as well as its Associated Countries. The consortium includes academia, businesses and law enforcement agencies (LEAs). The project brings data mining and predictive analytics tools to the next level, developing novel natural language processing (NLP), semantic multimedia analysis (SMA), social network analysis (SNA), Complex Event Processing (CEP) and artificial intelligence (AI) technologies. These technologies are combined for the first time and validated by 6 LEAs to collect, process, visualize and store online data related to terrorist groups, allowing them to take coordinated action in real-time while preserving the privacy of citizens.

The RED-Alert project aims to mine at least seven social media channels for content, and support at least ten languages for analysis. Therefore a good quantity and variety of data is analysed by the project that requires complex and (semi-)automatic solution that can efficiently process the data within legal constraints. This implies improved accuracy and usability of tools within the context of data privacy, as well as extended

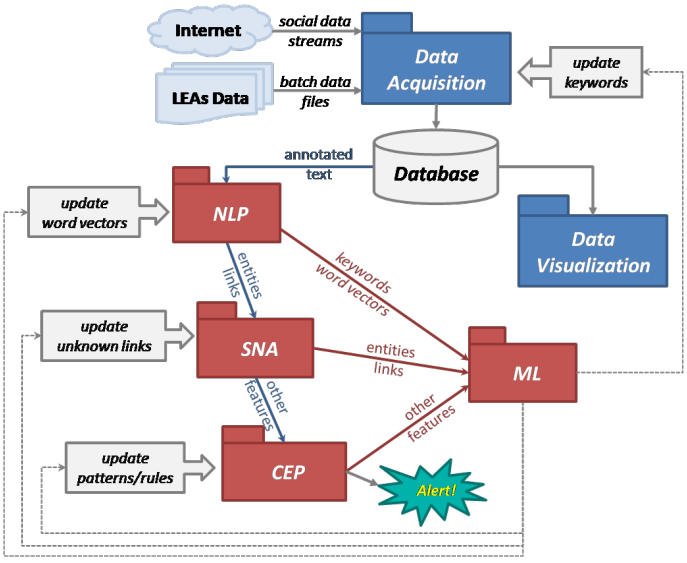

Fig. 1. High-level RED-Alert architecture

real-time and collaborative capabilities and support for further development.

In order to meet the project objectives, these software components (CEP, NLP, SMA, SNA, etc.) are developed by project partners. The integration of these software components brings innovation and impact in the everyday investigation of modern policing. The main challenge of the integrated solution is to assess social features in communications by terrorist organizations that will imply harmonisation of theories, tools and techniques from cognitive science, communications, computational linguistics, discourse processing, language studies and social psychology. Moreover, in order that the system performance to be adapted for each component the project implements a meta-learning process to assist each software component defined processes. A high-level architecture is shown in the Figure 1.

The other major challenge for the project is to preserve the privacy of citizens who use online social networking platforms. The project aims to enable secure and privacy preserving data processing; hence the malicious content and the corresponding personality can be tracked while the privacy of innocent citizens can be preserved. RED-Alert system is required to include privacy-preserving mechanisms allowing the capture, processing and storage of social media data in accordance to European and national legislations (especially GDPR [5]).

IV. Semantic Multimedia Analysis (SMA) Tool

Multimedia is extensively used in social networks nowadays and is gaining popularity among the users with the increasing growth in the network capacity, connectivity, and speed. Moreover, affordable prices of data plans, especially mobile data packages, have considerably increased the use of multimedia by different users including terrorists, who use social media platforms to promote their ideology and intimidate their adversaries. It is therefore very important to develop automated solutions to semantically analyse online multimedia contents.

The objective of the SMA Tool is to ensure security and policing of online contents by detecting terrorist material. It extracts meaningful information from multimedia contents taken from social media. The five main features of the tool are:

- Segmentation of audio streams, identifying sections of speech.
- Transcription of the segmented speech sections using an Automatic Speech Recognition (ASR) engine.
- Detection of sound events within audio streams, such as gunfire, explosions, crowd noise, etc.
- Extraction and identification of objects, such as logos, flags, weapons, faces, etc., within image and video scene elements.
- Extraction and transcription of text elements in image and video elements.

Moreover, the SMA Tool retrieves multimedia data, converts it to a uniform format and delivers the analysis results. The extraction of semantic information is the third of four stages the tool will perform. All four stages are as follows:

1) Input: Retrieval of multimedia files from disk or URL.
2) Stream Separation: Extraction of audio/video streams in multimedia files.
3) Feature Analysis: Semantic analysis of audio/image content.
4) Output: Compilation of results in a uniform JSON format.

A. Input options for the SMA tool

The SMA tool can process files which are stored locally or can optionally download media from a given URL. Currently the tool recognises the media format and type (audio/image/video) from its file extension, currently supported extensions are as follows:

- Audio File Extensions: .wav, .ogg, .mp3, .aiff, .flac
- Image File Extensions: .bmp, .jpeg, .jpg, .png, .tiff, .tif, .JPEG
- Video File Extensions: .avi, .mp4

B. Output format of the SMA tool

The SMA tool produces standard JSON formatted output for all types of multimedia input. A separate *.json* file is produced for each multimedia file processed. This file's name is simply the name of the original file with *.json* appended to it. For example, the output for the input file *Gun.jpg* would be called *Gun.jpg.json*. These output files are written into the directory from which the SMATool executable was run. All JSON output has two fields in common. These are:

- **file_name**: The name of the file that the analysis results apply to.
- **media_type**: The type of media the input file represents (*audio*, *image* or *video*).

JSON output for each of the supported media type have further fields which describe the results of the analysis which

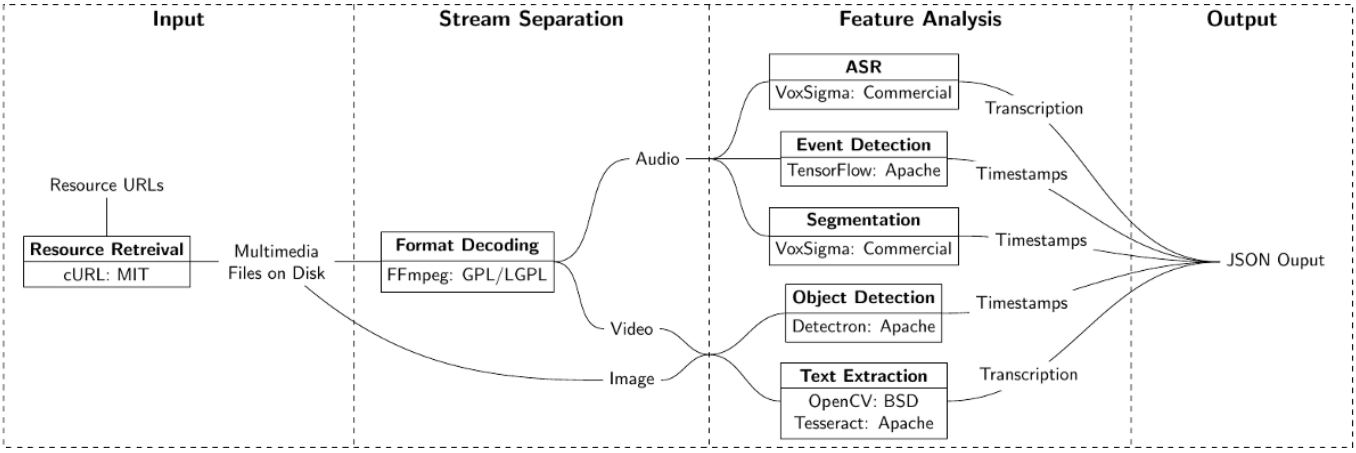

Fig. 2. SMA tool system diagram

apply to that type of media. The output is sent to the other key components of the project such as NLP, SNA and CEP. Component details of SMA Tool is provided hereunder.

C. Speech recognition

This component is used for audio segmentation, language detection and speech transcription. The RED-Alert project is required to support 10 languages, and be able to run offline, without having to send data to a 3rd party web API. We have consulted our LEA partners to prepare a list of 10 languages which must be supported by the speech / written text transcription elements of the SMA Tool. These languages are: Arabic, English, French, German, Hebrew, Romanian, Russian, Spanish, Turkish, and Ukrainian.

D. Face detection

The SMA tool uses a Haar-like feature based cascade classifier [6] to detect both frontal facing and profile faces in images. Haar-like features are calculated by finding the difference in average pixel intensity between two or more adjacent rectangular regions of an image. In the SMA tool, Haar cascades are used as a supplementary feature to implement simple face detection. More advanced techniques are implemented in the object detection element, which can also be used to detect people/faces.

E. Object detection

State of the art methods for detection of objects within images use large neural networks consisting of multiple subnetworks (region proposal network, classification network etc.). The SMA tools object detection utility uses the Faster R-CNN structure [7]. Faster R-CNN is constructed primarily of two separate networks: a Region Proposal Network (RPN) which produces suggestions of regions of an image which might contain objects, and a typical Convolutional Neural Network (CNN) which generates a feature map and classifies the objects in the proposed regions.

F. Audio event detection

Audio event detection is implemented in the SMA Tool by using a recurrent convolutional neural network [8]. The convolutional element classifies the short term temporal / spectral features of the audio, while the recurrent element detects longer term temporal changes in the signal. The SMA Tool apply feature extraction prior to processing by the network. This provides a more detailed representation of the audio signal to the network, meaning the first few layers can extract more meaningful information. Peak picking algorithms [9] are applied to remove any noise and only annotate the onset of any detected audio events.

V. Social Network Analysis (SNA)

In the last decades, human communication has gone through a crucial transition. Thanks to the Internet, which connects all individuals around the globe, everybody can contact each other without any time delay and without geographical restrictions. Social interactions became cheap and worldwide, the only restriction remained at the human side: all of us are able to process information at a finite rate and can engage trustful relations only with a few tens or hundreds of others. Therefore, describing and modelling of the new type of human interactions called for a description which is free of space limitations: these represent the tools of Network Science.

SNA module, aims to provide methods and software solutions for handling relational data. It focuses on three aspects of networked analysis as described in the following subsections.

A. Network dynamics and temporal network structure models

The tool describes the evolution of networks and edges/nodes in time, by calculating quantitative features derived from models on evolving networks, and evolution of communities.

Real systems evolve with time and are usually not static [10]. This can manifest in the emergence of new parts, the disappearance of existing parts, and also the relations among constituents can be rearranged over time. Temporal

networks with changing topology over time result typically changing community structures. Since community finding methods determine the structures only at different time steps, the structures from consecutive steps must be matched. When communities simply shrink or increase in size, then the matching is straightforward: matching of communities is determined uniquely by intersecting nodes between the two communities of different time steps. However, individuals can also change their community membership over time.

The SNA module implements a special community finder algorithm to solve this challenge. The solution is based on the property of the applied algorithm, which ensures, that adding new nodes and edges to a network does not change the membership status of a node or an edge. The only possible change is, that distinct communities fuse. This property allows an algorithm to match consecutive groups by introducing an intermediate time step, where the two snapshots are merged into a common network. Because the intermediate snapshot can contain only additional nodes and edges, the communities of the intermediate network can be matched to the prior and to the subsequent communities by the rule of matching intersections.

B. Link prediction solution

The SNA tool of the RED-Alert solution adopts network theoretic similarity and distance measures for counter terrorism purposes. Based on the special targeted measures, missing links and nodes are predicted by the module. Furthermore, some features e.g. weights, labels, directionality of the links are updated as well. The implementation relies on two theoretic pillars:

- prediction based on topological measures;
- prediction based on attribute information.

Topological measures use only information from connectivity patterns, in contrast attribute measures predict missing/hidden relations from common attribute statistics. Upon request of the analyst on the user interface of the integrated RED-Alert solution, the SNA module can apply hybrid predictions as well, where both networked measures and attribute data are combined [11].

It must be noted though, that all theoretical speculations are useless without reliable data sources. The scientific background behind this tool ensures only the mathematical rigour with the calculations, but the final conclusions must be always thoroughly reviewed by human experts. All mathematical models work with assumptions that can be only partially valid in real scenarios.

C. Hierarchy reconstructing methods

Terrorism has its own frame and structure. As all organizations that consist of many individuals and conduct several tasks, a hierarchical background drives the actions of terrorists. However, in several cases, this hierarchy is hidden and builds up in a self-organized way. For traditional observation techniques, this organization seems to be wide spread, unstructured and loosely connected. Here comes SNA into an important role: collecting small pieces of information from huge amount of data results in a holistic picture, where if data allows it the unseen hierarchical skeleton can be revealed.

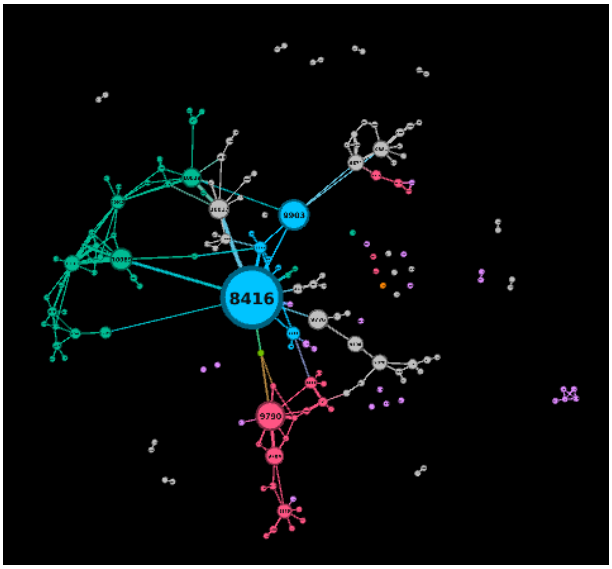

Fig. 3. Output of the SNA module

Here, algorithms are implemented for revealing hierarchical structures from flat dataset [12]. New networks are constructed from input data: either from co-occurrence statistics or from directed networks containing loops. Furthermore, quantitative measures are calculated for characterizing the similarity of any network to an ideal hierarchical structure [13].

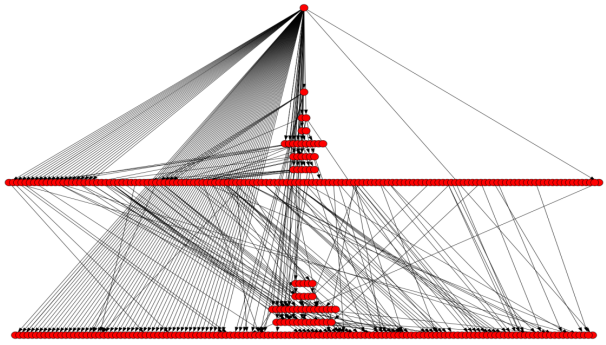

Fig. 4. Output of the SNA module

The first figure presents a typical thread-network layout of a forum in the Darkweb. The thread IDs are shown within nodes and the size of the nodes is proportional to the edges belonging to the given node. Node colours indicate topic groups; links are coloured by the dominant neighbouring node. The figure on the right shows the hierarchical structure of commenters of a Darkweb forum.

VI. Complex Event Processing (CEP)

The proposed solution of RED-Alert counts with a component specifically dedicated to the analysis of the post-processed social media publications and, then, trigger the alerts matching

a given pattern. Thus, the basic aim of this component is to identify, via pattern matching algorithms, the dynamics, interactions, feedback loops, causal connections and trends associated with the data content it receives as input from the other project components.

The CEP component aims to identify, via pattern matching algorithms, the dynamics, interactions, feedback loops, causal connections and trends associated with the data content it receives as input from the other RED-Alert components. Specifically, it is a secondary, downstream consumer of pre-processed data from the NLP, SNA and AI components and will generate output alerts; the component will also allow the configuration of data sources to allow the ingestion of external data out with the primary sources. The alerts themselves will be output to log files which will be monitored by a file reader component to display alerts, as well as monitor the CEP engine as a whole, and to integrate with the external APIs of the LEAs.

Specifically, it consumes the data from the NLP, SNA and AI components and will generate output alerts integrated with the RED-Alert dashboard so that the LEAs can trigger their investigative actions. Initially, the alerts will be by exact matches found in the patterns, but as the AI module is learning from previous patterns and the probabilistic values are being calculated and published downstream from the NLP, SMA and SNA components to the CEP, these patterns will increase its complexity. This probabilistic architecture is implemented to cope with the uncertainty associated with the data. The introduction of machine learning and AI will allow the CEP to improve its accuracy of the results and fully leverage the pattern matching capability.

As a development timeline, it comprises a data ingestion component which will, via connection components, acquire processed data from the previous components. Such ETL (Extract, Transform, Load) strategy has been implemented across two streams, one stream representing a live ingestion feed, the other a batch feed which will be used to analyse the data from a historical perspective over a configurable temporal range a configuration delivered by the configuration component. Ultimately, both streams are fed into the complex event processing component which itself will consist of multiple child components, each implementing a particular use case in response to specific LEAs requirements and investigative cases needs.

This block diagram shows the workflow, interactions, input/output and decision-making processes on the CEP engine itself. The engine itself works on structured, well defined JSON, where well defined includes all field names, their data types as well as an indication of their original source Note, in this case, source indicates where the data analytics (i.e. NLP, SNA and SMA processing) that generated particular aspects of the JSON originated, as opposed to the source of the input, i.e. the raw data.

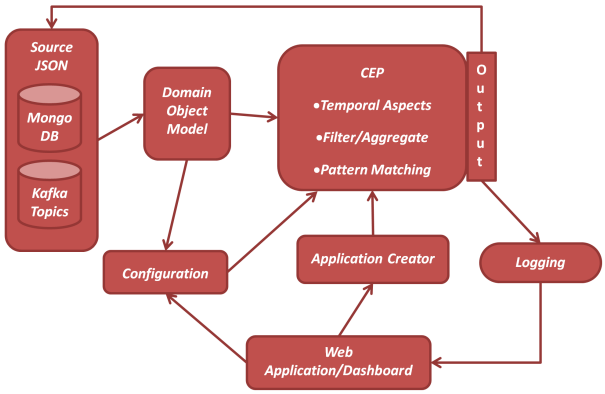

Fig. 5. CEP - Logical Component Diagram

VII. Discussions

Project RED-Alert is an ambitious project with significant impact in the modern policing of the online activities notably in predicting any imminent terrorist activity. The backend of the system requires smart and high-performance elements to ensure real-time delivery of the results. Semantic Multimedia Analysis (SMA) Tool, Social Network Analysis (SNA) Tool and Complex Event Processing (CEP) are the key components of the project solution that extracts information from a sheer volume of the input data. This is like finding a needle in a haystack. The reliability of the tools results require fine-tuned object detection techniques otherwise recurrent execution of the searches will have higher computing cost and performance overheads.

Besides technical challenges and project constraints, we have the obligation to address legal and ethical issues related to data collection, processing and storage. Data anonymization and visualisation [14] components positioning and features are the key elements to ensure compliance with the data protection legislations notably GDPR. However, the overall challenge of the project remains its integrated architecture performance and overall compliance with the legal requirements. The end users LEAs have the flexibility to adapt the features to their operational requirements and tune the components to match their needs.

VIII. Conclusions and Perspectives

We have presented our work of developing an online content analysis solution to improve security and policing of online activities and enable the law enforcement agencies to tackle imminent terrorist activities by analysing social media files without compromising privacy of the users of social platforms. Although RED-Alert project is of sensitive nature. However, the research leading to the project results could be applied in other fields to benefit from the impact of the integrated solution as well as integration of individual components into other systems.

The online contents analysis has several other commercial applications that can benefit from our work and leverage their competitiveness and product/services quality. One of the major

challenges of such analysis is to reduce the number of false positives and false negatives. We need to make finer grained tuning of RED-Alert components by using larger dataset of a broad range of objects and audio variations. Nowadays data collection, processing and storage have become itself very challenging due to the recently enforced GDPR compliance requirements. The situation is improving with the development of new data management processes and good practices for the data protection. We aim to further improve the performance of our work and evolve it towards a comprehensive online analysis. Another challenge to be addressed is to develop tools for hierarchical visualization of time evolving networks, which helps the analyst in understanding the possible correlations and trends at different scales.

ACKNOWLEDGMENT

The research leading to the results presented in this paper has received funding from European Commission Horizon 2020 (H2020) Programme under Research and Innovation Action H2020-SEC-12-FCT-2016 (Technologies for prevention, investigation, and mitigation in the context of fight against crime and terrorism). Grant agreement number 740688.

REFERENCES

[1] Chris Bousquet, Mining Social Media Data for Policing, the Ethical Way, online article of Gov. Tech, 2018 http://www.govtech.com/public-safety/Mining-Social-Media-Data-for-Policing-the-Ethical-Way.html

[2] Mohammad Tayebi, Uwe Glasser, Social Network Analysis in Predictive Policing: Concepts, Models and Methods, Springer Lecture Notes in Social Networks, 2016

[3] Policing 4.0: Deciding the Future of Policing in the UK, Deloitte report https://www2.deloitte.com/uk/en/pages/public-sector/articles/the-future-of-policing.html

[4] RED-Alert Project: http://redalertproject.eu

[5] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance)

[6] P. Viola and M. Jones, Repid Object Detection using a Boosted Cascade of Simple Features, in Computer Vision and Pattern Recognition, 2001.

[7] S. Ren, K. He, R. Girshick and J. Sun, Faster R-CNN: Towards Real-Time Object Detection with Region proposal Networks, 2016.

[8] I. Goodfellow, Y. Bengio and A. Courville, Deep Learning, MIT Press, 2016.

[9] C. Southall, R. Stables and J. Hockman, Improving Peak-Picking Using Multiple Time-Step Loss Functions, in Proceedings of the 19th International Society for Music Information Retrieval Conference (ISMIR), 2018.

[10] N. Masuda and R. Lambiotte, A Guide to Temporal Networks, Singapore: World Scientific, 2016

[11] Soundarajan, S., Hopcroft, J.: Using community information to improve the precision of link prediction methods. In: Proc. of WWW (2012)

[12] Tibly G, Pollner P, Vicsek T, Palla G (2013) Extracting tag-hierarchies. PLoS One. 2013 Dec 31;8(12):e84133. doi: 10.1371/journal.pone.0084133. eCollection 2013.

[13] E. Mones, L. Vicsek and T. Vicsek, Hierarchy measure for complex networks, PLoS ONE, 2012.

[14] W. Asif, I. G. Ray, S. Tahir and R. Muttukrishnan, Privacy-preserving Anonymization with Restricted Search (PARS) on Social Network Data for Criminal Investigations, 2018.