FCT-01-2015 - Forensics topic 1: Tools and infrastructure for the extraction, fusion, exchange and analysis of big data including cyber-offenses generated data for forensic investigation

See all projects funded under this programme / topic

Specific challenge:

The availability of petabytes of on-line and off-line information being open to the public owned by the Law Enforcement Agencies (LEA), such as police forces and/or custom authorities or the result of the investigation of a (cyber-) offence, represents a valuable resource but also a management challenge. Access to huge amounts of data, structured (data-bases), unstructured (multilingual text, multimedia), semi-structured (HTML, XML, etc.), heterogeneous data collected by LEA sensors such as Video, Audio, GSM and GPS, all possibly obfuscated or anonymized, available locally or over private LEA owned/shared networks or over the Internet, can easily result in an information overload and represent a problem instead of a useful asset.

Scope:

Proposals under this topic should aim to provide solutions at and beyond the state-of-the-art in the areas of intelligent use and management of complex and large amount of data for the discovery of correlated evidences to support forensic investigation on one hand and for the operational and situational awareness of law enforcement agencies on the other. The problem of extracting, integrating, exchanging ,analysing and exploiting large complex, structured and unstructured (Natural Language Text, SMS, multimedia) heterogeneous data, as well as that of exploiting unstructured data (Natural Language Text, SMS) and adding intelligence (trends analysis, scenarios, etc.), has to be solved by means of at and beyond state-of-the-art technologies in the areas of Big Data, Data Analytics, Multimedia Analysis, Data Modelling, Data mining, Visualization, Intelligent User's Interfaces, Information Retrieval, Automatic Language Translation, Weak Signal Analysis, Ontologies, High Level Fusion Techniques for Context Awareness and Knowledge Representation. Digital intelligence capabilities should also enable smart pre-processing and filtering of sensor data and stored data in order to improve their reliability, accuracy, accessibility and transmission volume.

The scope of this topic is threefold:

Firstly, tools and platforms should be developed for sampling, analysing, evaluating, interpreting, reasoning over, and recording forensic evidence from big data with a view to achieve solid and court-proof forensic evidence that can be used during legal prosecution. Applications should provide certainty with respect to the time and location of multimedia content and tests for authenticity and integrity of digital identities. Platforms should also provide users with semi-interactive techniques for understanding and visualizing data, including interdisciplinary approaches based on common, possibly standardized, ontologies and the exploitation of automated reasoning, information retrieval, and filtering tools. Human and organisational factors like multilingualism/multiculturalism as well as other trans-border issues (different terminologies, legislations, procedures) must be properly addressed.

Secondly, tools and platforms should be developed to enable LEAs to store, process, analyse, share, and exchange large amounts of heterogeneous data, including data arising from various types of sensors, with the aim of improving operational and situational awareness more efficiently. Data exchange between LEA and network operators shall be standardized for fast and efficient processing. These should include applications which can provide early warning signs (e.g. predictions of future trends). Vendor locking has to be excluded. The development of a base line system for current and future end users should also be envisaged and the solution should follow Open Source concepts. This will enable transparency, and continuous maintenance and development after the end of the project. The software should provide fine-grained authorisation mechanisms to regulate data access. Support for logging and in general maintain the chain of custody is also required.

Thirdly, tools and platform should allow reaching a significant speed-up in the whole process of analysing (cyber) offenses. The main challenges are the automation of as many analysis steps as possible; the countering of the obfuscation used by the attacker. The finding of an efficient way to identify an attacker despite use of anonymisation, , performing automatic deep analysis of all data in the offense, and making optimal use of the capabilities of man and

performing automated deep analysis of all data in the cloud, and making optimal use of the capabilities of man and machine.

Proposals addressing this topic should address the three aspects of the scope and take previous research at European and national level into account. Methodologies, standards, expertise and procedures for training, simulation, and testing investigations to empower the experts and stream-line the processes involved in the fusion, exchange and analysis of big data for forensic investigation and operational/situational awareness for law enforcement purposes should be considered.

The proposal will have to deal with the management of personal data, and related ethical and legal issues. Therefore considerable attention will have to be given to privacy and data protection, and to the adherence to European regulations. For each proposed solution, potential issues vis-à-vis these rights and regulations will have to be analysed, and recommendations on the best solutions to these issues must be proposed.

Proposals addressing this topic may involve the use of classified background information (EU or national) or the production of security sensitive results. As such, certain project deliverables may require security classification. The final decision on the classification of projects is subject to the security evaluation.

Proposals for this topic should take into account the existing EU and national projects in this field.

The Commission considers that proposals requesting a contribution from the EU of between €9m and €12m would allow this specific challenge to be addressed appropriately. Nonetheless, this does not preclude submission and selection of proposals requesting other amounts.

The outcome of the proposal is expected to lead to development from Technology Readiness Levels (TRL) 6 or above; please see part G of the General Annexes.

Expected impact:

Proposals should lead to:

- improved capabilities for the LEA to conduct investigations and analysis;
- higher efficiency in accessing relevant data sources and retrieving information significant for forensic investigation; and
- improved capabilities for trans-border LEA data-exchange and collaboration.

The outcome of the proposal is expected to lead to development from Technology Readiness Levels (TRL) 6 or above; please see part G of the General Annexes.

Type of action: Research & Innovation Actions

Record Number: 665096 / **Last updated on**: 2015-03-25

Retrieved on 2017-03-13

Permalink: http://cordis.europa.eu/programme/rcn/665096_en.html
© European Union, 2017