Introducing CopAS

CopAS tool

- fine-tuned production-ready framework running Elastic Platform developed in collaboration with Police CR (PCR)
- Bro, LogStash, ElasticSearch and Kibana
  - possible integration of other tools
- graphical user interface
- a set of pre-prepared dashboards and visualizations

- main emphasis on user-friendliness and ease of deployment & use
  - employs Docker for easier deployment
  - runs on all systems with Docker available (Windows, Linux, MacOS, ...)
- allows for generic usage (not only intended for PCR purposes)
Introducing CopAS
CopAS – container management

- copas ACTION [container name]
  - a tool for CopAS container management

```
[jeronimo@caine /home/jeronimo]$ copas -h
CopAS (Cops Analytic System) -- a system for data analyses using Elastic stack
* Created by Institute of Computer Science, Masaryk University, 2017 *

Usage: copas ACTION [container_name]
Available actions:
create ... creates a CopAS container (named 'container_name', if provided)
start ... starts a CopAS container (named 'container_name', if provided)
stop ... stops a CopAS container (named 'container_name', if provided)
destroy ... destroys a CopAS container (named 'container_name', if provided)
info ... shows information about available CopAS containers
monitor ... monitors the resource usage of CopAS containers
           (if -l|--live option provided, shows live resource usage)
enter ... enters a CopAS container (named 'container_name', if provided)
update ... updates the CopAS base image
           if a filename is provided, updates from the local image
```
CopAS – user environment

CopAS umožňuje nahrávat data různých podporovaných formátů do výkonné databáze ElasticSearch a tyto následně analyzovat a vizualizovat pomocí analytického nástroje Kibana.

Vyvíjeno Ústavem výpočetní techniky Masarykovy univerzity pod záštitou výpočetního Centra CERIT-SC (podpořeného projektem “CERIT Scientific Cloud”, LM2015085).
Ministry of Defence Research project

ANALYZA = Complex Analysis and Visualization of Large-scale Heterogeneous Data

– a research project submitted to the “Security Research Program of the CR for 2015-2020” of Ministry of Defence CR

– project goals: to develop a distributed system supporting complex analyses of heterogeneous data of large amounts
  – especially digital artifacts collected during police investigations

– the goal is to develop a system usable in 2021+
  • stable and scalable technologies
Basic Requirements I.

(ANALYZA = Complex Analysis and Visualization of Large-scale Heterogeneous Data)

The proposed/developed distributed system has to:

- deal with various **heterogeneous data**
  - network logs, financial logs, multimedia and document data, telecommunication data, real-world findings, ownerships, etc.
  - including large collections and/or larger data files
  - flexibility for future data types is a must

- allow **intra-domain as well as inter-domain analyses**
  - „Is there a community, which the subject regularly communicates with, no matter which technology is he/she using?“
  - inter-domain analyses performed in the same way as intra-domain ones

- allow **explorative (interactive) analyses**
  - analysts don’t know in advance, what they are looking for (the crime suspect is not always known)
  - the system has to allow for various types of queries and analyses
  - including local indications of suspects, evidences and findings
Basic Requirements II.

The proposed/developed distributed system has to:

– provide **useful and scalable views**
  – including visualizations of complex relationships
  – **generic** visualizations (graphs, location-based and time-based views, etc.) vs. **analysis-specific** visualizations

– support **collaborative team work**

– provide **high level of security**
  • even analysts from the same PCR team do not always share their data

– etc. etc.
Few Analyses Examples

(The ones that we implement as demonstration use-cases)

Smart Community Identification

– community of entities, which somehow cooperate on a crime
– can be identified over various data types (network and telecommunication communication, financial „communication“, known meetings, ...)

Suspicious Transactions Detections

– lookups using behaviour patterns
  • which can be used for different data types as well
– many research papers published detection methods of „money laundering“

Complex Network Analyses

– based on entity behaviour patterns
– currently deeply investigated using graph databases (Dgraph)
Few Analyses Examples

Pictures/Photos Analyses

- photos with 2 or more people (meetings)
- photos catching particular person
- children porn photos
- photos from particular environment (room)
- etc.

Location-based and/or Time-based entity behaviour

- based e.g. on GSM cell positions of travelling entities

And many many others ...

- PCR can provide lots of them
  - our demo use-cases are based on publicly available methods
Conclusions

Data analysis in cooperation with PCR

- interesting and attractive collaboration
  - parts of collaboration under NDA
  - many parts running under established mutual trust
  - *personal motivation*: building safer society 😊
- many open problems from various research areas
  - including artificial intelligence, natural language processing, etc.
  - etc.
  - colleagues/partners interested in such a collaboration still welcomed 😊
Thank you