Social Media Intelligence in Practice: The NEREUS Experimental Platform

Dimitris Gritzalis & Vasilis Stavrou

June 2015

Social Media Intelligence in Practice: The NEREUS Experimental Platform

NATIONAL CENTER FOR SCIENTIFIC RESEARCH "DEMOKRITOS"

3rd Hellenic Forum for Science, Technology & Innovation
Athens, June 2015

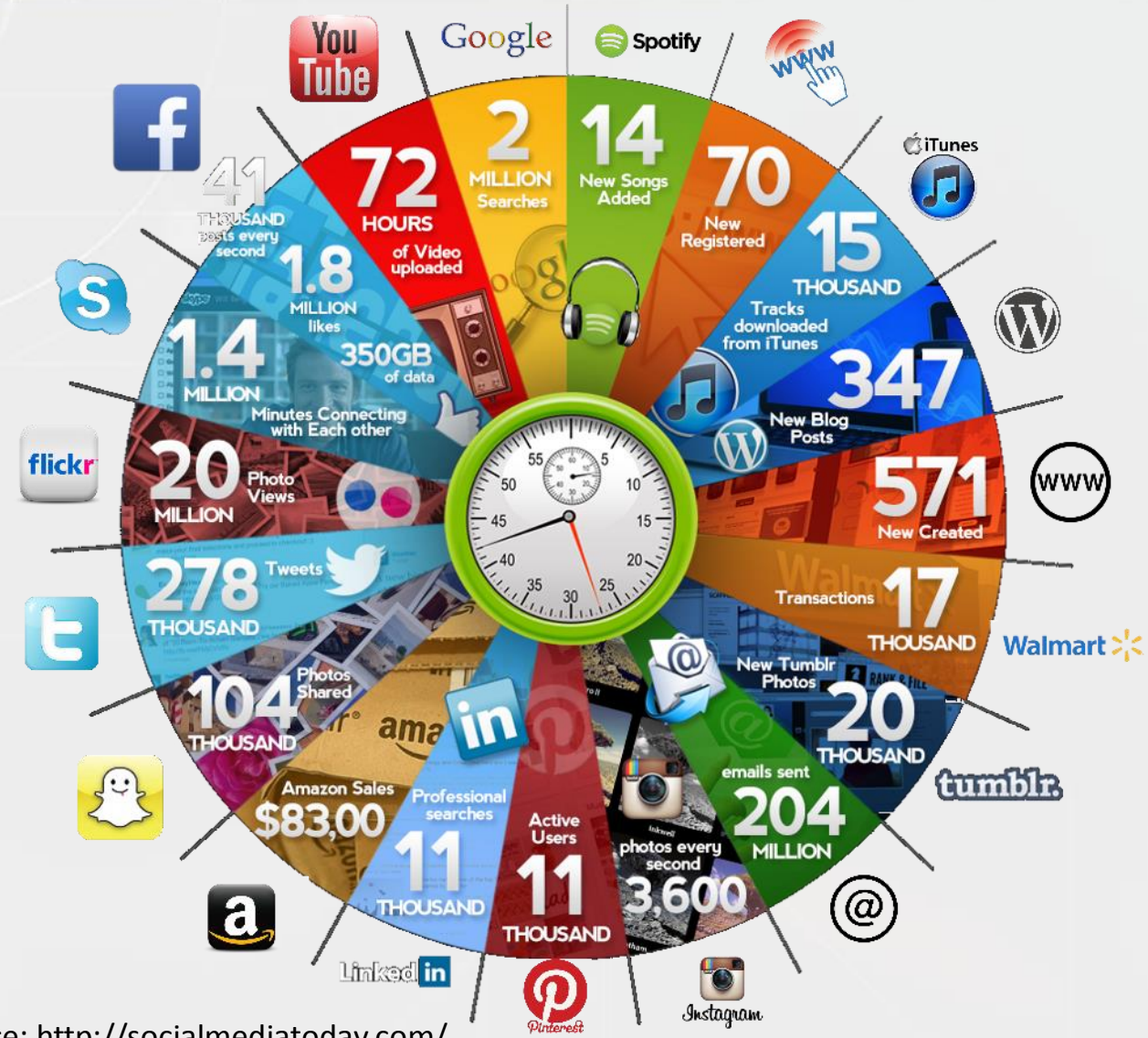
Dimitris Gritzalis & Vasilis Stavrou

Information Security & Critical Infrastructure Protection Laboratory
Dept. of Informatics | Athens University of Economics & Business

Presentation outline

- ⇒ Web 2.0 and Online Social Networks
- ⇒ Open Source and Social Media Intelligence
- ⇒ The NEREUS Framework
- ⇒ SOCMINT and behavior prediction capabilities
- ⇒ Conclusions

Web 2.0 and Online Social Networks (OSN)



Source: http://socialmediatoday.com/

Open Source & Social Media Intelligence

Open Source Intelligence (OSINT) is produced from **publicly available** information, which is:

- Collected, exploited and disseminated in a **timely** manner
- Offered to an **appropriate** audience
- Used for the purpose of addressing a specific **intelligence requirement**

Publicly available information refers to (not only):

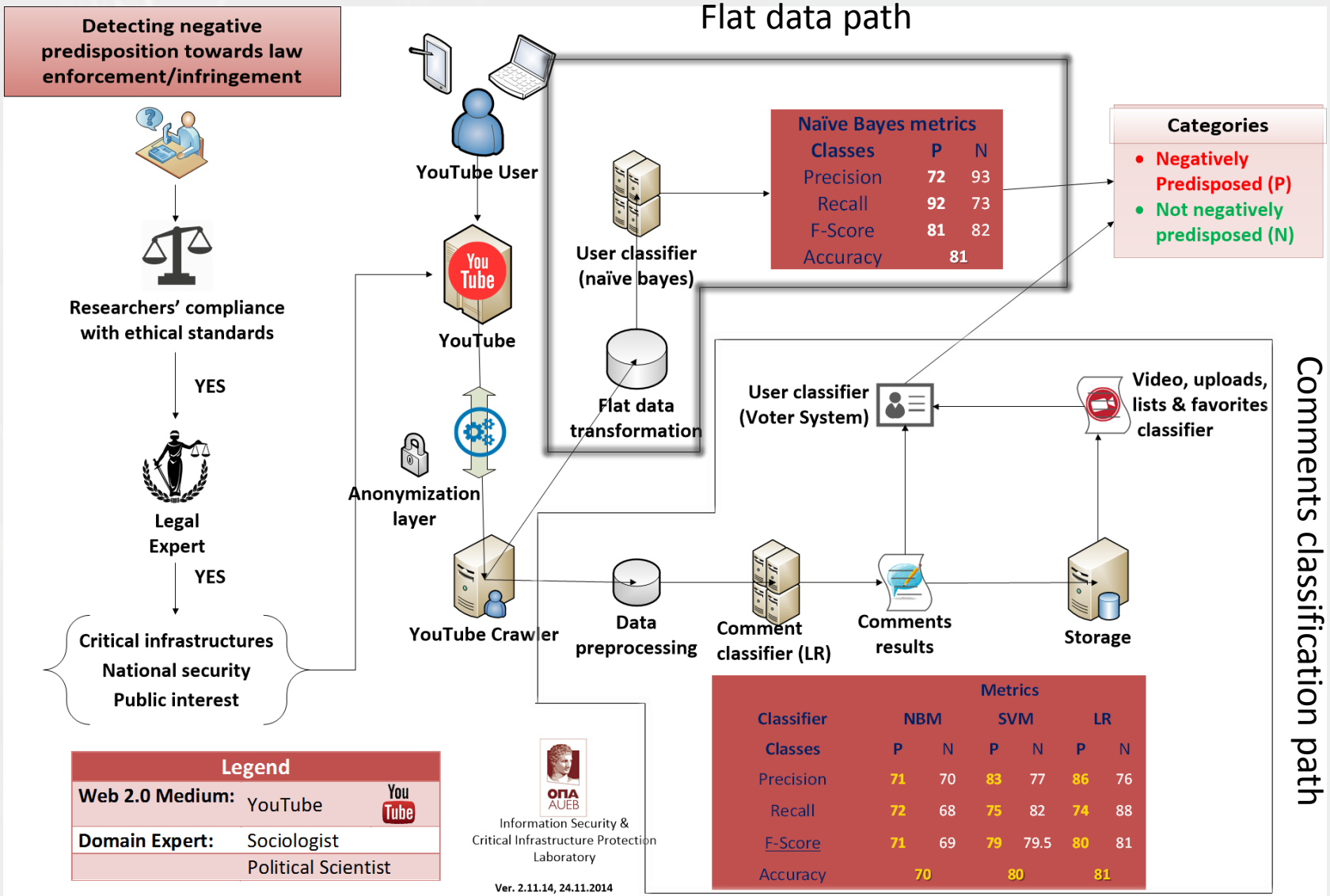
- **Traditional media** (e.g. television, newspapers, radio, magazines)
- **Web-based communities** (e.g. social networking sites, blogs)
- **Public data** (e.g. government reports, official data, public hearings)
- **Amateur observation/reporting** (e.g. amateur spotters, radio monitors)

Social Media Intelligence (SOCMINT) is produced from Online Social Networks and the Web 2.0

Revealing attitude towards law enforcement/infringement

OSINT	OSN: YouTube
Means utilized for the analysis	
Science	**Theory**
Computing	Machine Learning
	Data Mining
Sociology	Social Learning Theory
Applications: (a). Assist in detecting attitude towards law enforcement/ infringement (b). Assist in detecting deviant behavior of minors	

NEREUS: Architecture in a nutshell



The utmost importance of the social context

Authoritarian Regimes

Revealing personal attitude towards law enforcement/ infringement will be used by the Regime against resisting pro-civic rights movements.

Pro-civic rights movements should prevent such platforms from being used by the Regime, using any available means.

Democratic States

Revealing personal attitude towards law enforcement/ infringement may be used to protect Democracy from its opponents.

Democratic States may resist to social changes supported by, for example, grassroots political rights movements.

Democratic States may make use of such intrusive platforms, provided they are put under strict democratic control.

Revealing attitude towards law enforcement/infringement

Attitude towards law infringement

- Study: Motive, anger, frustrations, predisposition towards law enforcement/ infringement
- Means: Machine Learning, comment classification, flat data classification.

- ✓ Individuals tend to **transfer online** their offline behavior
- ✓ Identify users' **attitude towards law enforcement/infringement**
- ✓ Assist in detecting **delinquent behavior**
- ✓ Assist in predicting **deviant behavior of minors**

Dataset description

- Crawled YouTube and created dataset consists solely of **Greek** users.
- Utilized YouTube **REST-based API** (developers.google.com/youtube/):
  - Only publicly available data collected
  - Quote limitations (posed by YouTube) were respected
- Collected data were classified into three categories:
  - User-related information (profile, uploaded videos, subscriptions, favorite videos, playlists)
  - Video-related information (license, # of likes, # of dislikes, category, tags)
  - Comment-related information (comment content, # of likes, # of dislikes)

2.043.362 comments | **207.377** videos | **12.964** users

- Time span of collected data covered 7 years (Nov 2005 - Oct 2012).
- A basic anonymisation layer added to the collected data:
  - MD5 hashes instead of usernames

Machine Learning (1/2)

- Comment classified into categories of interest:
  - Process performed as **text classification**
  - Machine trained with **text examples** and the **category** each one belongs to
  - Excessive support by **field expert** (Sociologist)
- Test set used to evaluate efficiency of resulting classifier:
  - Contains pre-labeled data fed to machine, labeled by field expert
  - Check if initial assigned label is equal to predicted one
  - Testing set labels assigned by field expert
- Most comments written in Greek/greeklish
- Convertion of greeklish text to Greek
- Categories of content defined:
  - Users with a **negative** attitude towards law enforcement (**P**redisposed negatively (P))
  - Users with a **not negative** attitude towards law enforcement (**N**ot-predisposed negatively (N))

Machine Learning (2/2)

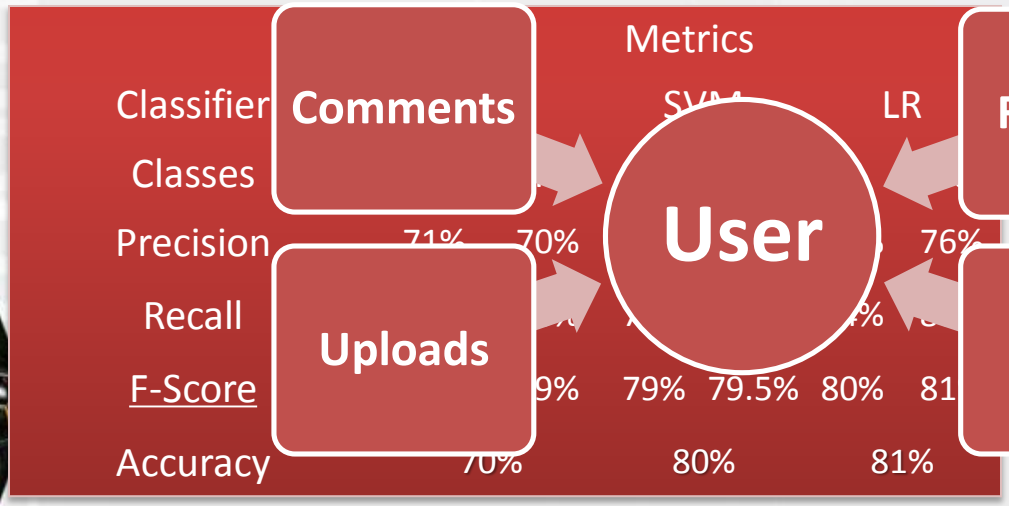
- **Video** classification:
  - Examination of a video on the basis of its comments
  - Voter process to determine category classification
- **(Video) Lists** classification:
  - Voter process to determine category classification (same threshold)
- Classifiers **efficiency** comparison:
  - Metrics (on % basis): Precision, Recall, F-Score, Accuracy
- Conclusions about **user behavior:**
  - If there is at least one category P attribute then the user is classified into category P
- **Logistic Regression** algorithm:
  - Best classifies a comment with **81% accuracy**

Precision: Measures the classifier exactness. Higher and lower precision means less and more false positive classifications, respectively.

Recall: Measures the classifier completeness. Higher and lower recall means less and more false negative classifications, respectively.

re: Weighted harmo-ean of both metrics.

Accuracy: No. of correct ifications performed by lassifier. Equals to the ent of good classifica- by all data.

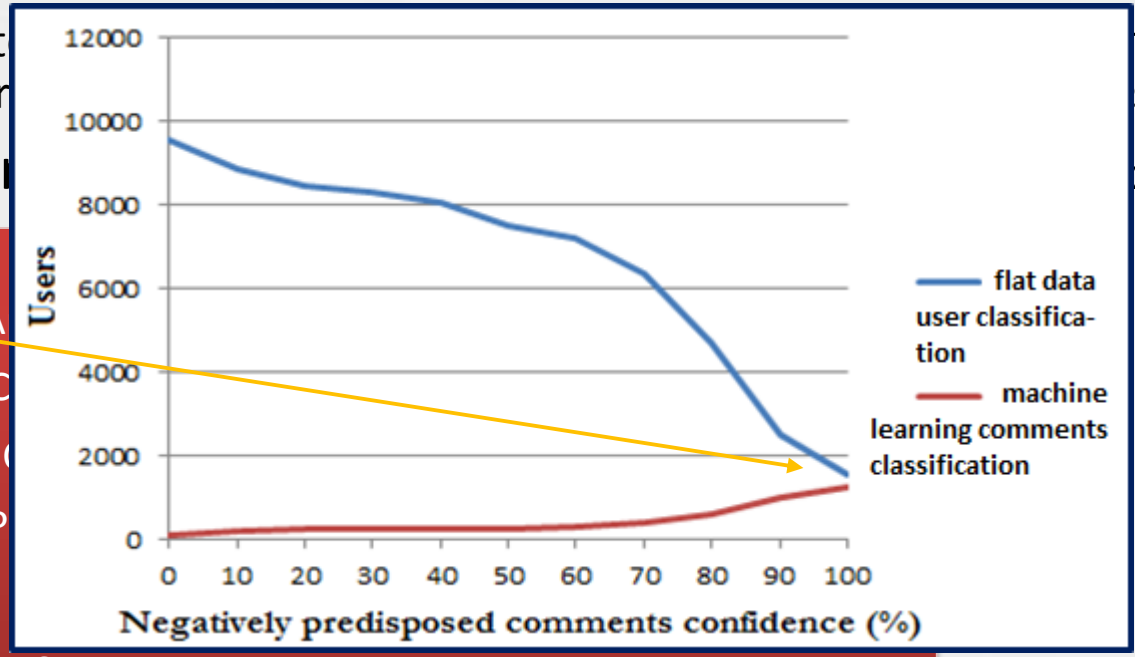


Analysis based on flat data

- Addressing the problem from a different perspective:
  - assumption-free and easy-to-scale method
  - verify (or not) the results of the Machine Learning approach
- Connection between users of category P and confidence of accuracy of comments belonging to category P.

Blue: Users of category P classified on the basis of the comment-oriented tuple (**Flat Data**).
Red: Users of category P classified on the basis of their comments-only (**Machine Learning**).

- Data transformation:
  - User represent… comment refer…
- Machine trai…



1721 users are (almost certainly) negatively predisposed towards law enforcement/infringement

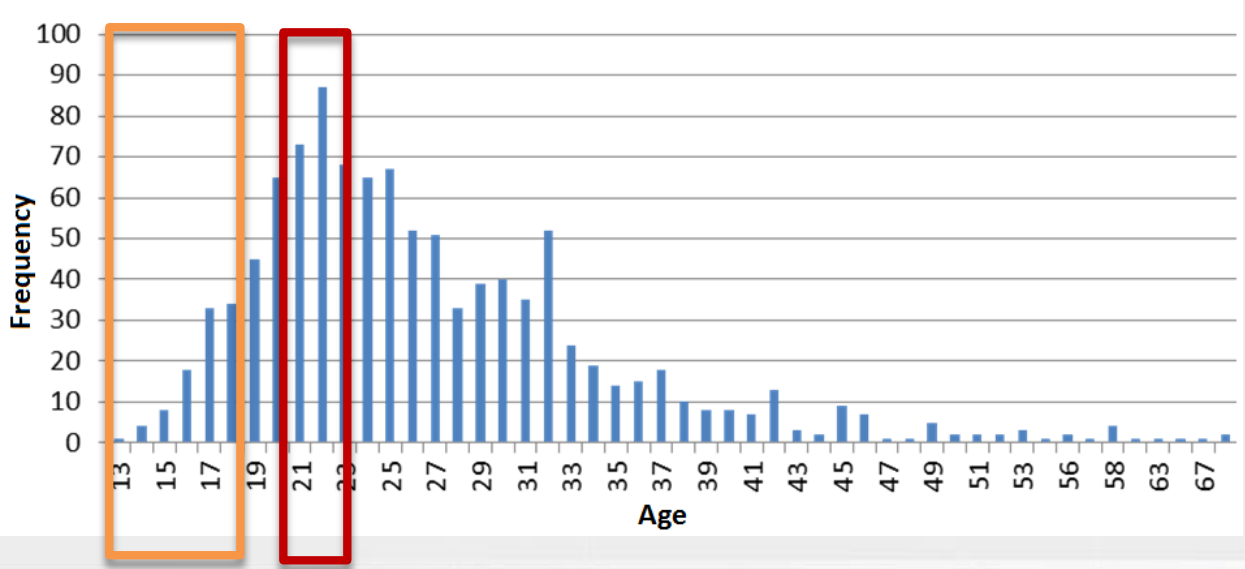
A…				
C…				
P…				
F-Score	80%	81%	81%	82%
Accuracy	81%		81%	

Selected observations

- ✓ **6%** of **comments** (among 2.000.000 collected) express **negative attitude** towards respecting the law (i.e., positive to law infringement)
- ✓ **3.5%** of **videos** (among 200.000 collected) classified into a specific category of interest
- ✓ **14%** of **users** (among 13.000 collected) express **negative attitude** towards respecting the law (i.e., positive to law infringement)

Ability to assist in **predicting delinquent behaviour** of **minors**

- Violent behaviour
- Cyber bullying
- Emotional or sexual harassment



General conclusions

- ✓ SOCMINT can transform into intelligence the vast amount of data produced by Web 2.0.
- ✓ SOCMINT is an intrusive technology and could put in danger civic rights.
- ✓ SOCMINT utilization is not - and should not be considered as - a solely technical issue.
- ✓ SOCMINT could assist in predicting attitude towards law infringement.
- ✓ SOCMINT could assist in predicting delinquent behavior of minors.

References

1. Gritzalis D., Stavrou V., Kandias M., Stergiopoulos G., “Insider Threat: Enhancing BPM through Social Media”, in *Proc. of the 6th IFIP International Conference on New Technologies, Mobility and Security* (NMTS-2014), Springer, UAE, 2014.
2. Gritzalis D., “Insider threat prevention through Open Source Intelligence based on Online Social Networks”, Keynote address, *13th European Conference on Cyber Warfare and Security* (ECCWS-2014), Greece, 2014.
3. Gritzalis D., Kandias M., Stavrou V., Mitrou L., "History of Information: The case of Privacy and Security in Social Media", in *Proc. of the History of Information Conference*, Law Library Publications, Athens, 2014.
4. Kandias M., Mitrou L., Stavrou V., Gritzalis D., “Which side are you on? A new Panopticon vs. privacy”, in *Proc. of the 10th International Conference on Security and Cryptography* (SECRYPT-2013), pp. 98-110, Iceland, 2013.
5. Kandias M., Galbogini K., Mitrou L., Gritzalis D., "Insiders trapped in the mirror reveal themselves in social media", in *Proc. of the 7th International Conference on Network and System Security* (NSS-2013), pp. 220-235, Springer (LNCS 7873), Spain, June 2013.
6. Kandias M., Virvilis N., Gritzalis D., "The Insider Threat in Cloud Computing", in *Proc. of the 6th International Conference on Critical Infrastructure Security (*CRITIS-2011), pp. 93-103, Springer (LNCS 6983), United Kingdom, 2013.
7. Kandias M., Stavrou V., Bozovic N., Mitrou L., Gritzalis D., "Can we trust this user? Predicting insider’s attitude via YouTube usage profiling", in *Proc. of 10th IEEE International Conference on Autonomic and Trusted Computing (*ATC-2013), pp. 347-354, IEEE Press, Italy, 2013.
8. Kandias M., Stavrou V., Bosovic N., Mitrou L., Gritzalis D., “Proactive insider threat detection through social media: The YouTube case”, in *Proc. of the 12th Workshop on Privacy in the Electronic Society* (WPES-2013), pp. 261-266, ACM Press, Germany, 2013.
9. Kandias M., Virvilis N., Gritzalis D., “The Insider Threat in Cloud Computing”, in *Proc. of the 6th International Workshop on Critical Infrastructure Secu-rity* (CRITIS-2011), Bologna S., et al (Eds.), pp. 93-103, Springer (LNCS 6983), Switzerland, 2011.
10. Kandias M., Mylonas A., Virvilis N., Theoharidou M., Gritzalis D., “An Insider Threat Prediction Model”, *in Proc. of the 7th International Conference on Trust, Privacy, and Security in Digital Business* (TrustBus-2010), pp. 26-37, Springer (LNCS-6264), Spain, 2010.
11. Mitrou L., Kandias M., Stavrou V., Gritzalis D., "Social media profiling: A Panopticon or Omniopticon tool?", in *Proc. of the 6th Conference of the Surveillance Studies Network,* Spain, 2014.
12. Stavrou V., Kandias M., Karoulas G., Gritzalis D., "Business Process Modeling for Insider threat monitoring and handling", in *Proc. of the 11th International Conference on Trust, Privacy & Security in Digital Business* (TRUSTBUS-2014), pp. 119-131, Springer (LNCS 8647), Germany, September 2014.