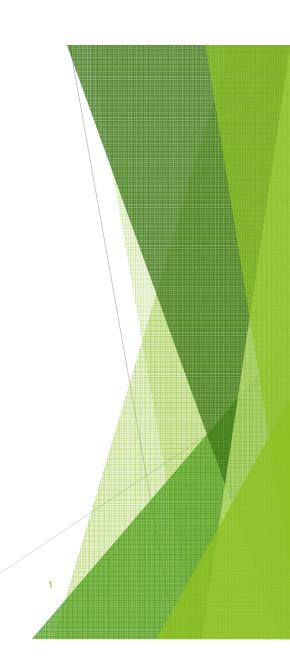
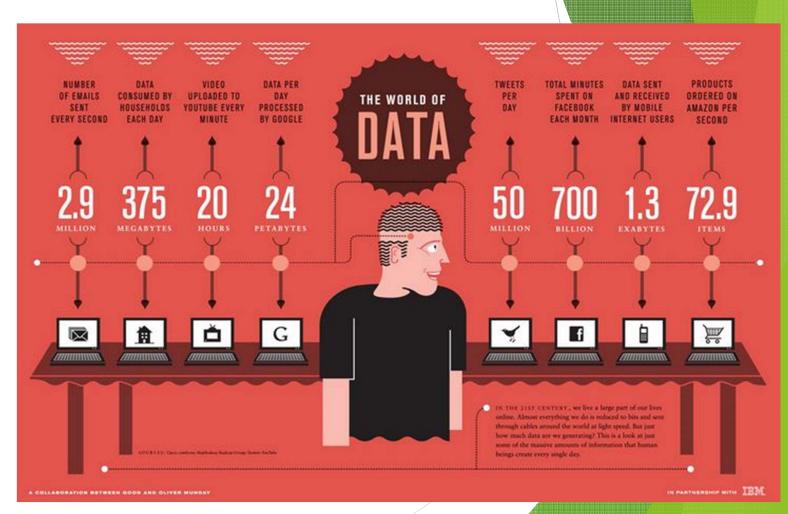
# Big Data and Cyber Security A bibliometric study

Jacky Akoka, Isabelle Comyn-Wattiau, Nabil Laoufi Workshop SCBC - 2015 (ER 2015)



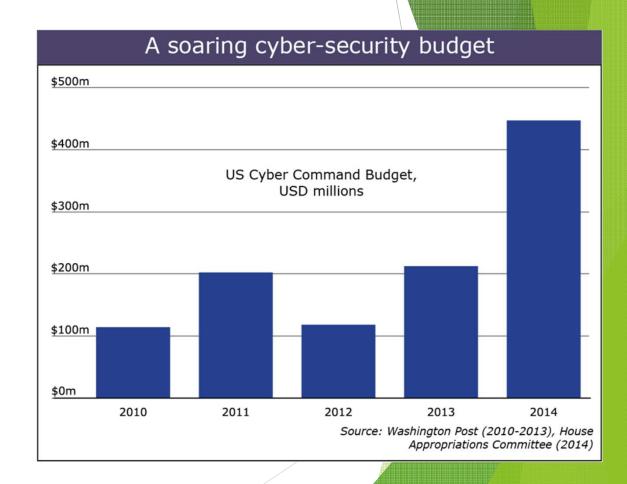
### **Big Data**

☐ a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling the highvelocity capture, discovery, and/or analysis (Gantz)



### Cyber-security

Cybersecurity refers to the techniques, processes and methodologies concerned with thwarting illegal or dishonest cyber attacks in order to protect one or more computers on any type of network from any type of damage.



## Cybersecurity and Big Data

Structured, analytical, repeatable

Security Intelligence Platform

#### Real-time Processing

- · Real-time data correlation
- · Anomaly detection
- · Event and flow normalization
- · Security context & enrichment
- · Distributed architecture



#### Security Operations

- · Pre-defined rules and reports
- Offense scoring & prioritization
- · Activity and event graphing
  - Compliance reporting
    - · Workflow management

Big Data Platform

**IBM Security** 

Intelligence

with Big Data

### **Big Data Processing**

- · Long-term, multi-PB storage
- · Unstructured and structured
- · Distributed Hadoop infrastructure
- · Preservation of raw data
- Enterprise Integration



#### Analytics and Forensics

- · Advanced visuals and interaction
- · Predictive & decision modeling
- · Ad hoc queries
- Interactive visualizations
- · Collaborative sharing tools
- · Pluggable, intuitive UI

Creative, exploratory, intuitive

## Examples of cyber attacks

- Spamming: sending unsolicited bulk messages to multiple recipients
  - ▶ By 2015, the spam volume is 95% of all email traffic
- Search Poisoning: dishonest use of Search Engine Optimization techniques to falsely improving the ranking of a webpage.
- Botnets: networks of malware-infected compromised computers managed by an adversary.
- Denial of Service (DoS): makes a system or any other network resource inaccessible to its intended users
- Phishing: fraudulently acquires confidential user data by mimicking ecommunication, mainly through email and web spoofing
  - Malware: software programmed to perform and propagate malicious activities
  - Website Threats: attackers exploiting vulnerabilities in legitimate website, infecting them and indirectly attacking visitors

## Our research approach

- Objective: a bibliometric study on research combining big data and cybersecurity
- Main steps:
  - Find the number of research articles using the terms Big Data and Cybersecurity in the abstract, and/or in the title, and/or in the keywords
  - Focus degrees of papers on the big data and cybersecurity issues (abstract, keywords, title)
  - Comparing number of publications by researchers and by professionals
  - Analyze the publications according to different perspectives:
    - ▶ Security component (threat, asset, vulnerability, control, security requirement, security treatment, event, risk, impact)
    - ► IT artifact (language, metamodel, system design, ontology, taxonomy, framework, architecture, methodology, guideline algorithm, method fragment, metric, prototype)
    - ▶ Application domain (military, healthcare, education, public sector, banking, tourism, marketing)
    - Objective (traceability, integrity, confidentiality, availability, privacy, reliability, scalability, performance, usability, quality, security)
    - Intervention level (network security, software security, information system security, security audit, security policy

# Results: Research papers vs. periodicals

Nb of papers Perspectives	with chain in A	with chain in K	with chain in T	With chain in AKT
Bigdata & cybersecurity	11	8	2	21
Security Component Artifact	4 2	0	2	7 2
Usage	4	0	0	4
Objective	4	1	1	6
Intervention level	0	0	0	0

Nb of papers	with	with	with	with
	chain	chain	chain	chain
	in	in	in	in
Perspectives	A	K	T	AKT
Bigdata &	25	0	2	27
cybersecurity				
Security Component	12	0	0	12
Artifact	0	0	0	0
Usage	7	0	0	7
Objective	12	0	0	12
Intervention level	0	0	0	0

Examples of periodical: Business Intelligence Journal, Channel Insider, etc.

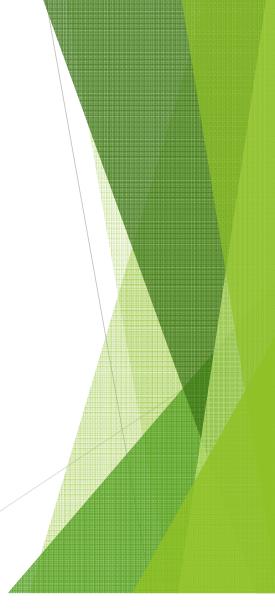


### Main findings

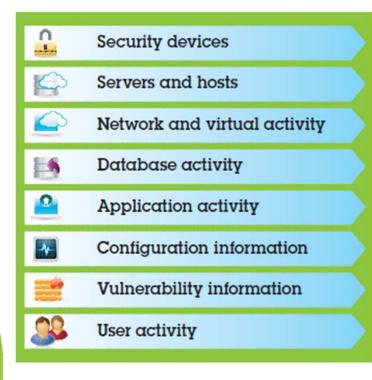
- Very few research publications on both topics (big data and cybersecurity)
- Security components addressed:
  - ▶ Risk, threat, vulnerability, event, control
- Quality objectives mentioned:
  - Security, privacy, performance, usability
- Artifacts produced:
  - ▶ Methodology, framework, algorithm, prototype
- Application domains addressed:
  - Healthcare
- Nothing on intervention level (audit, policy, requirements, etc.)
- That means large research opportunities!

# Big data and cybersecurity: Characterizing the intersection

- Cybersecurity for Big Data
  - Data in the cloud
    - ► Data leakage
  - Data analytics
    - ▶ Secure Hadoop and Map-Reduce (privacy): avoid record linkage
- Big Data for Cybersecurity
  - Social network analysis
  - Mobile data analysis
  - Security analytics
    - ▶ Detect anomalous behaviours in near real-time
    - ▶ Obtain clusters for typical malicious behaviors
    - ▶ Go back in time from security events to root causes
    - Discover suspicious applications



## Big data for security: an example (IBM)



#### **Event correlation**

- Logs
- Flows
- IP reputation
- Geographic location

### Activity baselining and anomaly detection

- User activity
- Database activity
- Application activity
- Network activity

### Offense identification

- Credibility
- Severity
- Relevance

High priority security offenses

Other stakeholders: Accenture, HP, Invensys, EADS, CISCO, Unisys, etc.

# Security for big data Top ten challenges (Cloud security alliance)

