Panel on Data and Multimedia

Services based on Sensing Data: Handling with Care Sensitive Data

Moderator

Andy Snow, Ohio University, USA

asnow@ohio.edu

Wednesday, April 25, 15:45 - 17:30

NexComm 2018, April 22-26, 2018 - Athens, Greece

Panelists

- Pascal Urien, Telecom ParisTech, France
- Yoshihisa Udagawa, Tokyo Polytechnic University, Japan
- Corneliu Octavian Dumitru, German Aerospace Center (DLR), Germany
- Jerzy Grzymala-Busse, University of Kansas, USA
- Jedrzej Rybicki, Forschungszentrum Juelich GmbH, Germany
- Keith Mayes, Royal Holloway, University of London UK

Sensor Service Challenges

- **Sensors generate lots of data – exponential growth from IoT expected**
- **TeraBytes passe, think ExaBytes and Zetabytes**
- **Heterogeneous data**
- **Services required for sensor data**
  - **Generic services -- industry sector agnostic**
  - **Specialized services – industry sector specific**
- **Technological advances necessary to handle**
  - **Quality of Services in the face of massive data volumes**
  - **Formidable Dependability issues**

Heterogeneous Data

- Text
- Video
- Audio
- Images
- Structured
- Unstructured

- Streaming
- Files

Sensor Service Needs

- Rich set of services
- Scalable services
- Responsive (Fast) services
- Ubiquitous services
- Economical services
- Dependable services

Sensor Service Dependability

- Safety
- Reliability
- Maintainability
- Availability
- Data Confidentiality
- Data Integrity
- Resiliency

Sensor Service Dependability

- Safety
- Reliability
- Maintainability
- Availability
- Data Confidentiality
- Data Integrity
- Resiliency

Security

Examples of New Technological Strategies

- Super-Servers? Intermediate Servers?
- Partitioned Servers?
- Caching?
- AI Driven Service Utilities and Search Algorithms
- Smart Data Summaries
- Pattern Recognition
- Scalable Architectures
- Computing Power

Panelists

- Pascal Urien, Telecom ParisTech, France
- Yoshihisa Udagawa, Tokyo Polytechnic University, Japan
- Corneliu Octavian Dumitru, German Aerospace Center (DLR), Germany
- Jerzy Grzymala-Busse, University of Kansas, USA
- Jedrzej Rybicki, Forschungszentrum Juelich GmbH, Germany
- Keith Mayes, Royal Holloway, University of London UK

Machine Learning in Cyber Security

Jerzy W. Grzymala-Busse

’ University of Kansas, Lawrence, KS 66045, USA

Machine Learning in Cyber Security

- Machine Learning in Cyber Security is used to find anomalies, prevent fraud and abuse
  - identifying malicious behavior
  - identifying malicious entities
    - hackers
    - attackers
    - malware
- Removing malware

Machine Learning in Cyber Security

- Google is using Machine Learning to prevent malware on smartphones using Android
- Amazon uses a Machine Learning based service for S3 cloud storage
- Analysts of ABI Research estimate that machine learning in cyber security will boost spending to $96 billion to 2021

Modern Vehicles, Driving Across the Cyber-Physical Divide

...a few personal opinions...

Professor Keith Mayes

Director of the Information Security Group

ROYAL HOLLOWAY UNIVERSITY OF LONDON

The Cyber-Physical Divide

Cyber (IT) systems historically have focussed on information processing, with human computer interfaces being the main touch-points to the real world.

- *Physical sensing, access and control solutions have long existed, but generally were isolated/closed and limited functionality systems.*

Today, many devices are becoming Internet enabled and this is merging cyber systems with the physical world; creating the Internet of Things (IoT)

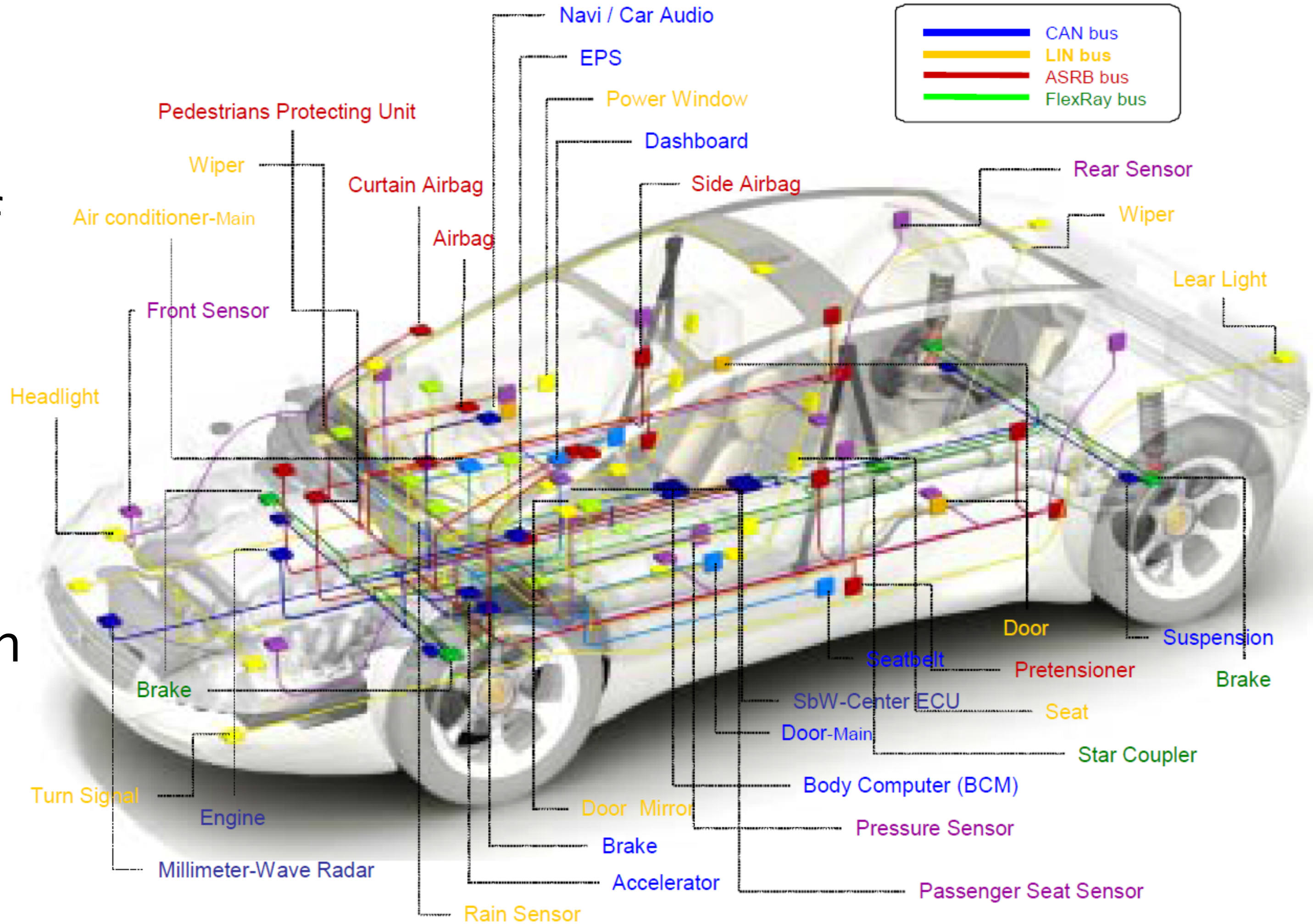
- *Companies are rushing towards interesting products and services, but without enough care for security*

A cyber-attack could now led to physical damage and serious safety risks

Driving Across the Cyber-Physical Divide

Modern Cars

- 100+ CPUs
- 100M lines of source code
- Sensors, controls and safety critical algorithms
- Mixed up with infotainment and mobile comms



New Cars are Safer – Right?

By conventional measures (crash tests) new cars are safer than older ones, but what about after an IT security/safety MOT?

- *We have had a fatality in a driverless car, due to an "algorithm" rather than security attack*
- *We have attacks against various car systems, plenty of targets!*

Safety:	**Privacy:**
Start/stop engine control	Multiple cameras
Electric hand brakes/Anti-collision breaking	Phone contacts/calls via car systems
Park assist	GPS tracking
Air bag trigger	Sat-Nav History
Auto lights	
Keyless entry	

Thank you!

www.linkedin.com/in/keithmayes

www.royalholloway.ac.uk/isg


Pascal.Urien@telecom-paristech.fr
ICDT 2018
Athens
April 25th 2018
Blockchain
The Next Big Thing ?
Towards Blockchain IoT (BIot)

Illustration of a BIoT Ethereum Transaction

Overview | Comments

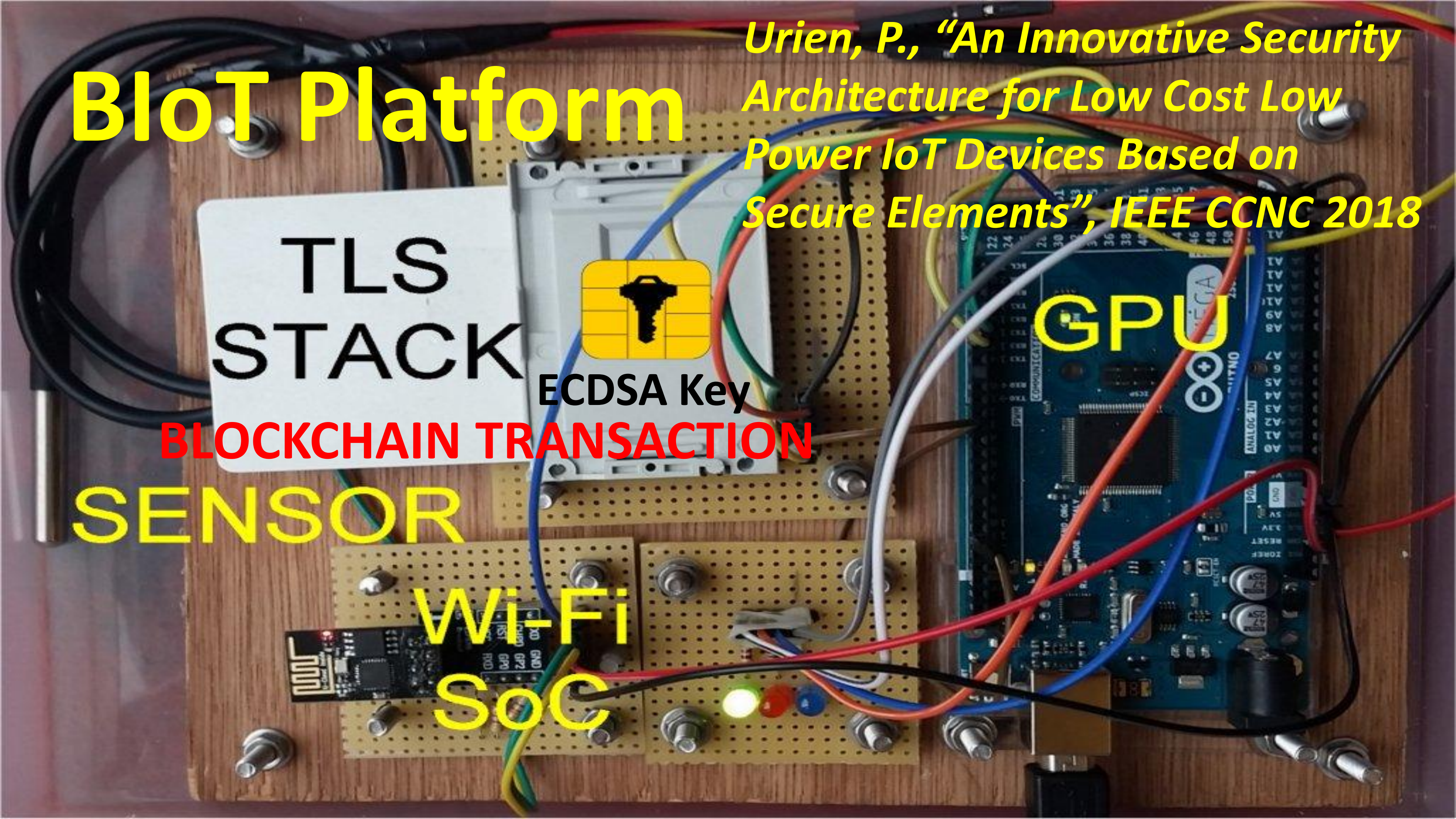
Transaction Information

Tools & Utilities

https://etherscan.io/tx/0xd6904d832462ae17718c69e9caa0c3f3bed458382ac1f4e43b1aadd8e94744ad

TxHash:	0xd6904d832462ae17718c69e9caa0c3f3bed458382ac1f4e43b1aadd8e94744ad
TxReceipt Status:	Success
Block Height:	4942834 (561272 block confirmations)
TimeStamp:	94 days 18 hrs ago (Jan-20-2018 09:52:42 PM +UTC)
From:	0x6bac1b75185d9051af740ab909f81c71bbb221a6
To:	0x6bac1b75185d9051af740ab909f81c71bbb221a6
Value:	0 Ether ($0.00)
Gas Limit:	80000
Gas Used By Txn:	22020
Gas Price:	0.00000003 Ether (30 Gwei)
Actual Tx Cost/Fee:	0.0006606 Ether ($0.41)
Nonce:	12
Input Data:	0xTemperature=25C

Switch Back


BIoT Platform
Urien, P., "An Innovative Security Architecture for Low Cost Low Power IoT Devices Based on Secure Elements", IEEE CCNC 2018
TLS STACK
ECDSA Key
BLOCKCHAIN TRANSACTION
GPU
SENSOR
Wi-Fi SoC

Ethereum Transaction

```
F8 74 // RLP List, length= 116 bytes
   0C // nonce 1 byte =12 decimal
   85 06FC23AC00 // gasPrice = 30 GWei
   83 013880     // gasLimit = 80000 gas
   // recipient address 20 bytes
   94 6BAC1B75185D9051AF740AB909F81C71BBB221A6
   80 // Null Ether Value
   // Data 15 bytes "Temperature=25C"
   8F 54656D706572617475726553D323543
   1B // recovery parameter, 1 byte (27=+, 28=-)
   A0 // r, 32 bytes, ECDSA r parameter
   A9B58980F76EE6284800B82A2B5DF13E456887EC0CF426A5E5D6A738EB1784ED
   A0 // s, 32 bytes, ECDSA s parameter
   629633C6A3ED5FEE0FB40E2D1CF251345B885D372857B1A6C4762C9BE914281F
```

Public key is recovered from the signature two solutions + (27) and (-) (28)

https://etherscan.io/tx/0xd6904d832462ae17718c69e9caa0c3f3bed458382ac1f4e43b1aadd8e94744ad

Blockchain: a public ledger

Bitcoin 12LZjvQBy31ABRpqvMZQbu7S9K5SxaifjW in Block 96188

Full Bitcoin Block 96188

Share:

block, address, transaction Search

PREVIOUS BLOCK

NEXT BLOCK

Number Of Transactions	4
Output Total	91.24 BTC
Height	96188
Time	2010-12-07 13:57:40
Difficulty	8,078.19525793
Bits	453516498
Version	1
Nonce	944968343
Block Reward	50 BTC
Days Destroyed	2

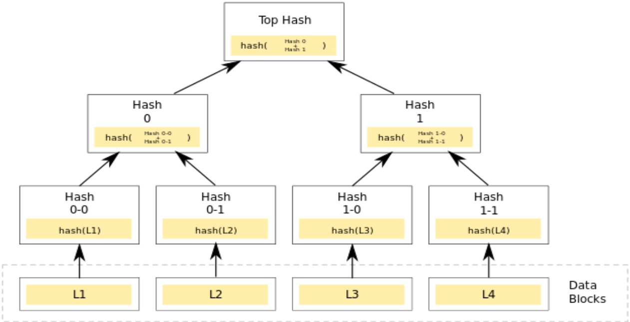
PoW

Hash	000000000004d6e22b42bf66fd9cad1977bdee13abab1e51a2d03ca22e6f71af
Previous Block	0000000000027c094bf08f7c27a6debcb5f36419bf53390e3e1c40a653e2195c
Next Block(s)	000000000042c9d08f3f06d502a247f090625fb6c7623cf956a38e987c59e0f
Merkle Root	26ab6b697b8e06416b2fe5047f558c997af0a9c36e6a6eae79ff59745b5065a1

tx:d4a73f51ab7ee7acb4cf0505d1fab34661666c461488e58ec30281e2becd93e2 33.59 BTC Fee: 0 BTC

prev tx 1689LPUuixaxSchENLMNaNbS3hYVgdpaSS -33.59 BTC		13RoCeq4K8ddPW6ugcheFoXK4GC2BLVuET 0.05 BTC	next tx
		12LZjvQBy31ABRpqvMZQbu7S9K5SxaifjW 33.54 BTC	next tx

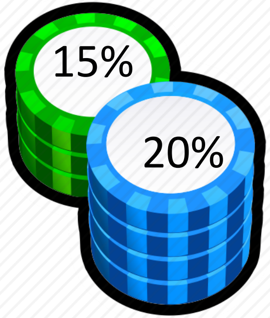
Transaction identifiers are stored in a Merkle Tree



The Magic of PoW: Mining

- H(x): sha256((sha256)(x))
- Solve: H(nonce, header) < (65535 << 208) / Difficulty
  - Given the **computation difficulty D**, and the **hashrate (h(t), in computations per second)**, the probability Δp of solving the PoW in Δt second is
    - $\Delta p = \Delta t\, h(t)/D$
  - The mining duration follows an exponential distribution, whose probability density function ρ(t) is :
    - $\rho(t) = \lambda e^{-\lambda t}$ with $\lambda = h(t)/D$ in s-1

hi/h is a token



- h= Σ hi, the computation power is shared by miners
  - The probability to win the mining process is hi/h

Questions ?

Blockchain Transaction Protocol for Constraint Nodes

draft-urien-core-blockchain-transaction-protocol-00.txt

https://tools.ietf.org/html/draft-urien-core-blockchain-transaction-protocol-00

Earth Observation Big Data: New Paradigms

Corneliu-Octavian Dumitru, Gottfried Schwarz, Mihai Datcu

DLR

The particular challenges

Challenge 1: Volume and heterogeneity

Challenge 2: Big EO Data Analytics

Challenge 3: Big EO Data Mining

Challenge 4: Human Machine Communication

Challenge 5: Information platform

Challenge 1: Volume and heterogeneity

- EO images: multisensory, eg. MS, SAR, altimeter, etc.
- These are multidimensional signals, acquired by sensors or instruments
- Sensor data carry physical meaning, radiation level, wavelength, etc.
- They are measuring land, ocean, or atmospheric parameters
- The VHR EO images observe detailed spatial structures and objects
- Satellite Image Time Series observe evolution processes over long period of time.

- An important particularity of EO images should be considered, is their “**instrument**” nature, i.e. they are **sensing physical parameters, and they are often sensing outside of the visual spectrum.**

- All these are autonomous sources with distributed and decentralized control.
- In this context Big EO Data seeks to explore complex and evolving Earth processes their inter-relationships impacting environmental, socio – economic phenomena.
- Therefore, Big EO Data has another very specific dimension, the large and diverse areas of applications and users in a meta-diversity of landscape of disciplines.

Challenge 2: Big EO Data Analytics

- The today techniques, methods, and tools, for automated data analysis are insufficient for the analysis and information extraction from EO data sources.

- A new goal has become the gathering of the user's interest, together with the transformation of the data into reduced information and knowledge items, and adaptation to direct and easy understanding.

- The capability of retrieving information interactively and the use of data-driven paradigms are now more than ever necessary due to the huge data volumes being involved.

Challenge 2: Big EO Data Analytics

Methods of **Computer Vision** and **Pattern Recognition**, are needed for new tasks:

- **Detecting, localizing** and recognizing objects
- Recognition and extraction of **semantic descriptions** of the scenes from sensor data
- Extract **quantitative measures** of the physical meaningful parameters of the scene
- **Registration** of multi-sensor multi-temporal data
- **Exploit variability of the imaging modes** to provide different types of information about various structures
- Recognition methods to **distinguish huge variability of scene classes** and objects with very good precision

DLR

Challenge 3: Big EO Data Mining

Big data involves more and more **machine or statistical learning** for “**discovery**” functions

The discrepancy between **data volume** explosion and **analysis potential** is continuously growing, new solutions are required:

- Detection of **irrelevant** data

- New sensors as based on **Compressive Sensing/Sampling**, recoding smaller data volumes but with the pertinent content
- Data **compression**

- **Machine/statistical learning** algorithms for fast prediction
- **DNN** for large scale prediction
- Content analysis to extract **higher-level analytics**
- Extraction and **formalization of knowledge** for data classification and understanding

DLR

Challenge 4: Human Machine Communication

Predictive, adaptive natural User Interfaces

Learning and anticipating the user **behavior** and collaborate with the user.

Understand and learn the user **intentions and context**, establish a **dialog**

Transform **non-visual sensor** data and information in human easy understandable representations.

Challenge 5: Information platform

Web based interactive technologies and tools

Distributed architecture systems

To cope with very important load and requirements regarding the data volumes to be accessed, the complexity of the information to be extracted, analyzed and presented, the adaptations to specific applications, and speed of interactive operation.

Cloud computing should enable tasks not achievable with actual resources.

But, new methods are further needed, since tools as Hadoop or MapReduce reached their limits.

Potential solutions are foreseen in virtual EO data center frames connected and communicating across clouds for enhanced potential to share hardware resources and data.

SERVICES BASED ON SENSING DATA:
Handling with Care Sensitive Data

25. April 2018 | Jedrzej Rybicki | Juelich Supercomputing Center

SENSITIVE RESEARCH DATA

Research (Data):

- Open (Reproducibility)
- Provenance
- Replication

SENSITIVE RESEARCH DATA

Research (Data):

- Open (Reproducibility)
- Provenance
- Replication

Examples:

- Medical Trails
- p-Medicine
- Brain Scans
- Language Recordings

BEST PRACTICES IN SENSITIVE DATA MANAGEMENT

Workflow:

1. get consent

BEST PRACTICES IN SENSITIVE DATA MANAGEMENT

Workflow:

1. get consent
2. protect identities (either anonymization or not collecting personal data)

BEST PRACTICES IN SENSITIVE DATA MANAGEMENT

Workflow:

1. get consent
2. protect identities (either anonymization or not collecting personal data)
3. regulate access (user group, usage time)

BEST PRACTICES IN SENSITIVE DATA MANAGEMENT

Workflow:

1. get consent
2. protect identities (either anonymization or not collecting personal data)
3. regulate access (user group, usage time)
4. store personal data separately and securely (e.g. anonymization log)

BEST PRACTICES IN SENSITIVE DATA MANAGEMENT

Workflow:

1. get consent
2. protect identities (either anonymization or not collecting personal data)
3. regulate access (user group, usage time)
4. store personal data separately and securely (e.g. anonymization log)

Anonymization (EC Opinion 05/2014):

- is it possible to single out individuals
- is it possible to link records belonging to an individual
- can information be inferred concerning an individual

SENSITIVE DATA AND SENSITIVE RESEARCH

Conflicts:

- Researcher-Community
- Researcher-Subject
- Subject-Subject
- Researcher-Society

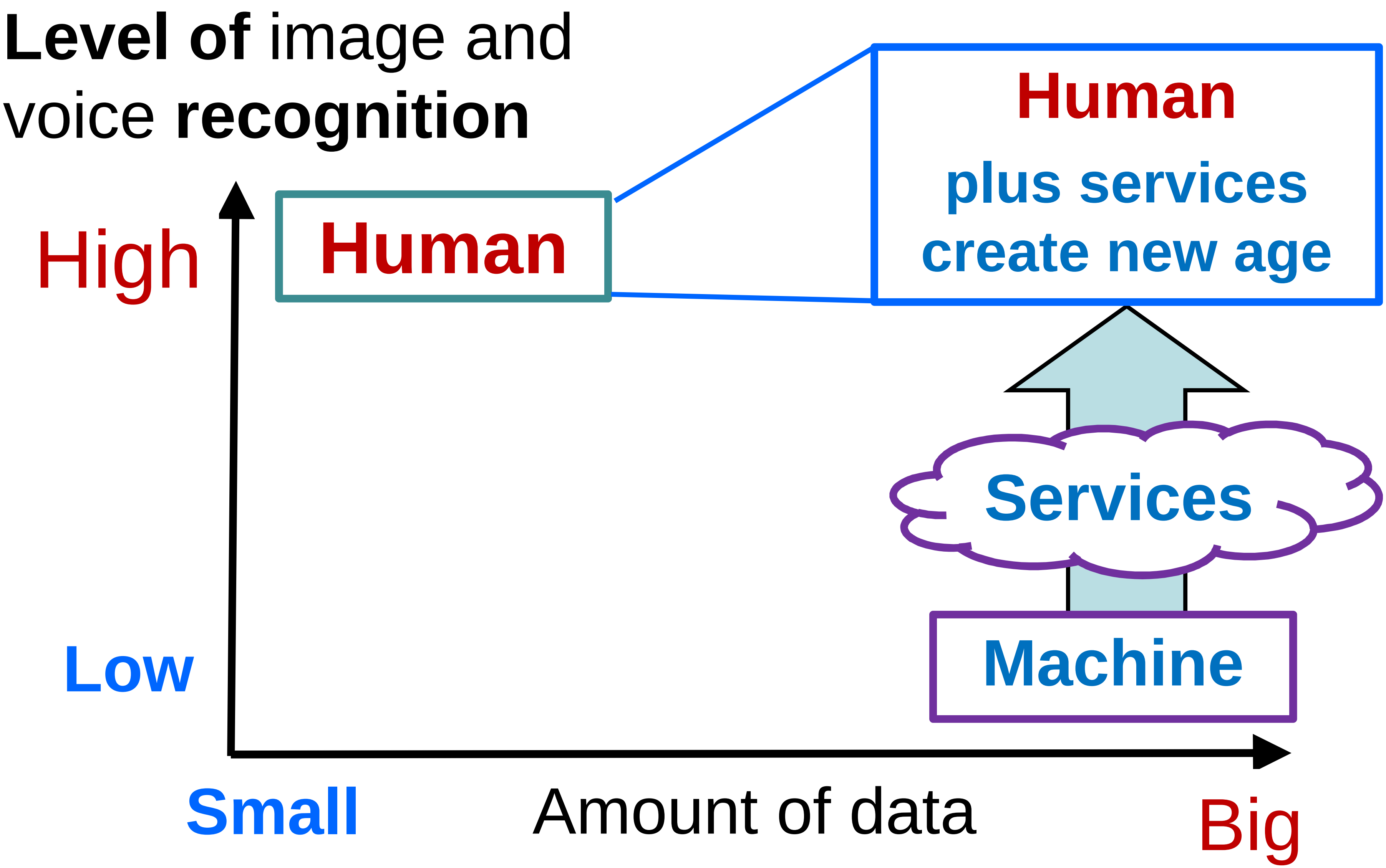
SENSITIVE DATA AND SENSITIVE RESEARCH

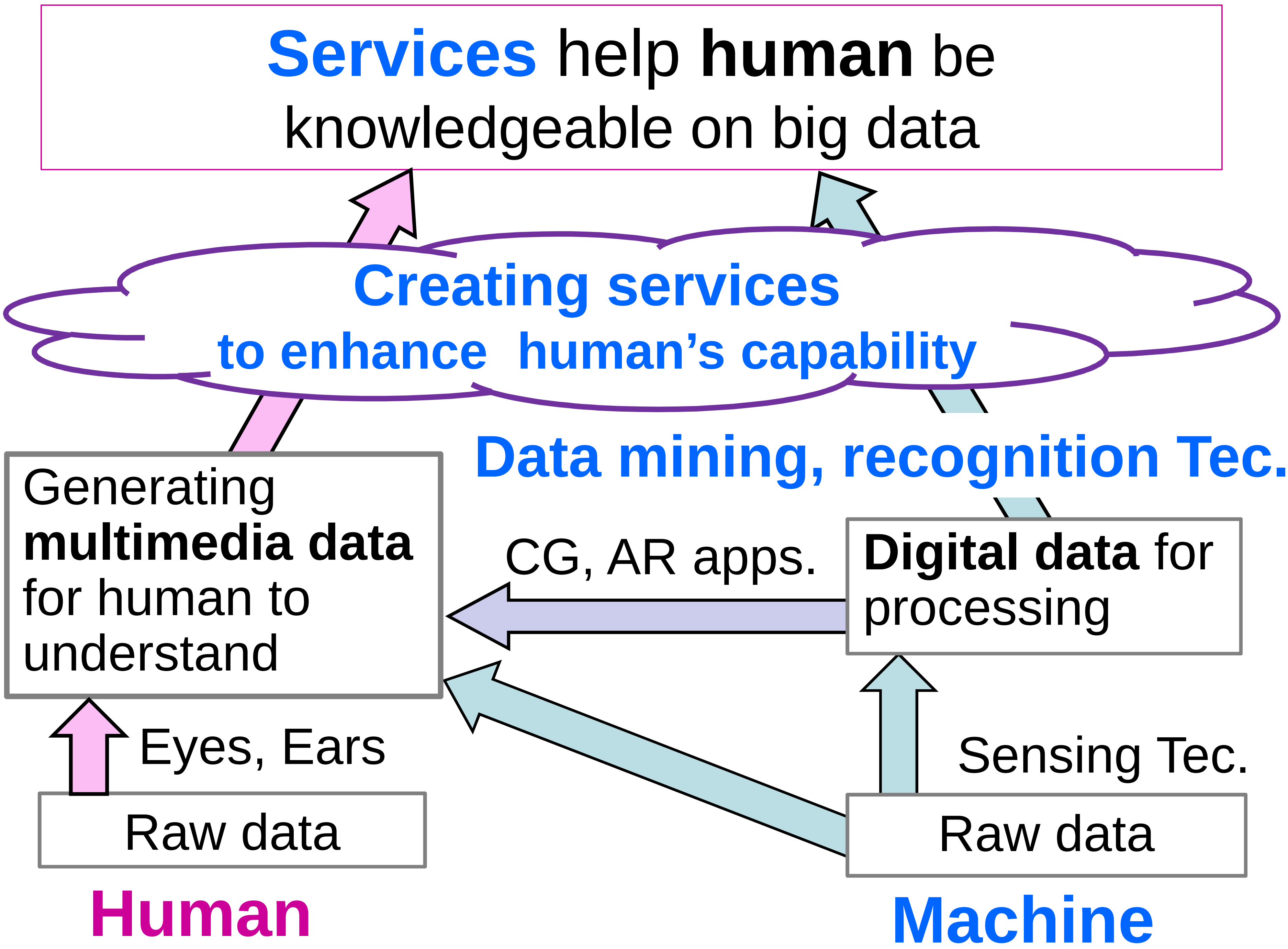
Conflicts:

- Researcher-Community
- Researcher-Subject
- Subject-Subject
- Researcher-Society

How to conduct responsible research?

Creating Services for Human
Based on Sensing Data





Services help human be knowledgeable on big data
Creating services
to enhance human's capability
Data mining, recognition Tec.
Generating multimedia data for human to understand
CG, AR apps.
Digital data for processing
Eyes, Ears
Sensing Tec.
Raw data
Raw data
Human
Machine

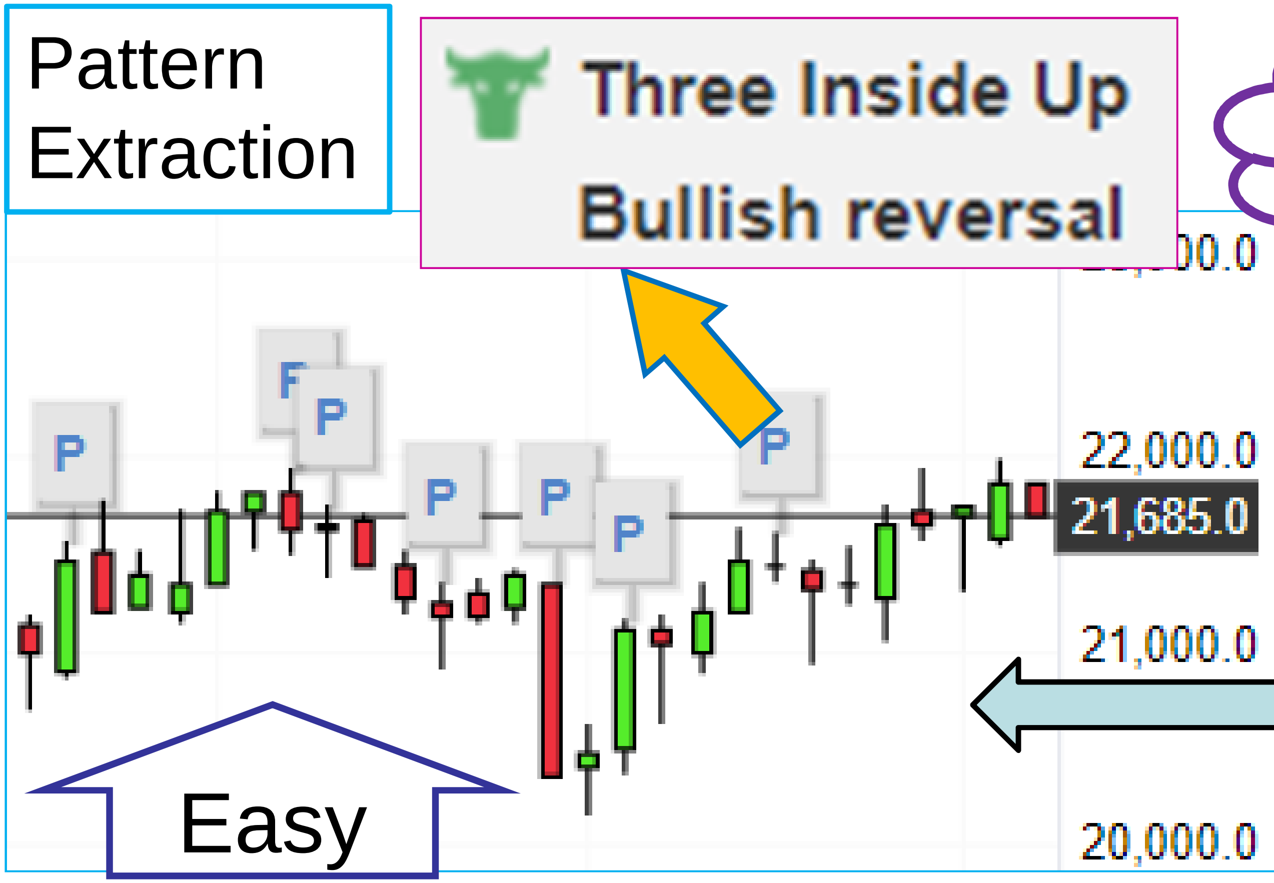
The interested topics

1. Multimedia needs and services to help human's recognition.
    - image data,
    - fintec,
    - health care, etc.
2. Are multimedia data essential or supplementary in your application?
3. Purpose of use of multimedia data:

    recognition ⇒ understanding
    ⇒ prediction ⇒ decision making, etc.

Case of Stock Price (fintec)

Pattern Extraction



Reliability Is low

Multimedia data

Date	Open	High	Low	Close*	Vol(M)
10-Apr-18	24,198	24,511	24,199	24,408	368
9-Apr-18	24,037	24,373	23,955	23,979	380
6-Apr-18	24,373	24,434	23,738	23,933	395
5-Apr-18	24,313	24,622	24,314	24,505	347

Raw data