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
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
Cyber (IT) systems historically have focused 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 lead to physical damage and serious safety risks**

Information Security Group
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!*

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

Information Security Group
Thank you!

www.linkedin.com/in/keithmayes
www.royalholloway.ac.uk/isg
Blockchain: The Next Big Thing? 
Towards Blockchain IoT (BIot)
**Transaction Information**

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TxHash</strong></td>
<td>0xd6904d832462ae17718c69e9caa0c3f3bed458382ac1f4e43b1aadd8e94744ad</td>
</tr>
<tr>
<td><strong>TxReceipt Status</strong></td>
<td>Success</td>
</tr>
<tr>
<td><strong>Block Height</strong></td>
<td>4942934 (561272 block confirmations)</td>
</tr>
<tr>
<td><strong>TimeStamp</strong></td>
<td>94 days 18 hrs ago (Jan-20-2018 09:52:42 PM +UTC)</td>
</tr>
<tr>
<td><strong>From</strong></td>
<td>0x6bac1b75185d9051af740ab909f81c71bba221a6</td>
</tr>
<tr>
<td><strong>To</strong></td>
<td>0x6bac1b75185d9051af740ab909f81c71bba221a6</td>
</tr>
<tr>
<td><strong>Value</strong></td>
<td>0 Ether ($0.00)</td>
</tr>
<tr>
<td><strong>Gas Limit</strong></td>
<td>80000</td>
</tr>
<tr>
<td><strong>Gas Used By Txn</strong></td>
<td>22020</td>
</tr>
<tr>
<td><strong>Gas Price</strong></td>
<td>0.000000003 Ether (30 Gwei)</td>
</tr>
<tr>
<td><strong>Actual Tx Cost/Fee</strong></td>
<td>0.0006606 Ether ($0.41)</td>
</tr>
<tr>
<td><strong>Nonce</strong></td>
<td>12</td>
</tr>
<tr>
<td><strong>Input Data</strong></td>
<td>0xTemperature=25C</td>
</tr>
</tbody>
</table>

Illustration of a BIoT Ethereum Transaction

https://etherscan.io/tx/0xd6904d832462ae17718c69e9caa0c3f3bed458382ac1f4e43b1aadd8e94744ad
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 54656D70657261747572653D323543
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

PoW

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
    - \( Δp = Δt \frac{h(t)}{D} \)
  - The mining duration follows an exponential distribution, whose probability density function \( ρ(t) \) is:
    - \( ρ(t) = λe^{-λt} \) with \( λ = \frac{h(t)}{D} \) in s\(^{-1}\)
- \( h= Σ h_i \), the computation power is shared by miners
  - The probability to win the mining process is \( \frac{h_i}{h} \)
Questions?

Blockchain Transaction Protocol for Constraint Nodes
draft-urien-core-blockchain-transaction-protocol-00.txt
Earth Observation Big Data: New Paradigms

Corneliu-Octavian Dumitru, Gottfried Schwarz, Mihai Datcu
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
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
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

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
BEST PRACTICES IN SENSITIVE DATA MANAGEMENT

Workflow:

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

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
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)

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
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

Level of image and voice recognition

High

Human

Human plus services create new age

Low

Small

Amount of data

Big

Services

Machine
Services help human be knowledgeable on big data

Creating services to enhance human’s capability

Generating **multimedia data** for human to understand

**Human**

- Eyes, Ears
- Raw data

**Machine**

- Digital data for processing
- Raw data
- Sensing Tec.

Data mining, recognition Tec.

CG, AR apps.
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 $\Rightarrow$ understanding
     $\Rightarrow$ prediction $\Rightarrow$ decision making, etc.
Reliability is low

Case of Stock Price (fintec)

Pattern Extraction

Three Inside Up
Bullish reversal

Easy

Multimedia data

Raw data

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