The Industrial Internet Revolution and Digital Transformation Challenges.

HỘI NGHỊ KHOA HỌC QUỐC GIA LẦN THỨ XII
NGHIÊN CỨU CƠ BẢN VÀ ỨNG DỤNG CÔNG NGHỆ THÔNG TIN

Keynote FAIR, Hue, Vietnam 07/06/2019

Doan B. Hoang, Prof.
School of School of Electrical and Data Engineering
VICS-Virtualized Infrastructures and Cyber Security
University of Technology Sydney, Australia
Introduction
- The connected world of Internet of Things
- Industrial Internet: what is it? Why embrace it? why now?

I4.0 Technologies
- Address the underpinning technologies that make Industrial Internet possible

Challenges and Opportunities
- Where do we go from here?
Moving from Internet of Things to Industry 4.0
The Internet of Things (IoT) is an information network of IoT objects (sensors, machines, robots, cars, buildings, homes, cities, data, processes, etc.) that integrates and allows interaction and cooperation among them to reach common objectives (e.g., self-driving cars).
The IoT application covers “smart” environments/spaces in domains such as: Transportation, Building, City, Lifestyle, Retail, Agriculture, Factory, Supply chain, Emergency, Health care, User interaction, Culture and tourism, Environment and Energy.
Industrial Revolution

1969 Internet – only 50 years ago
2007 First smartphone : iPhone - only 12 years ago
2009 First definition of Cloud computing from UC Berkeley - 10 years ago
2011 Internet of Things
2014 Industrie 4.0 (Germany)

Cyber-Physical Systems for Agricultural and Construction Machinery Current Applications and Future Potential
Georg Jacobs, Felix Schlüter, Jan Schröter, Achim Feldermann
In the same way the Internet created enormous value by interconnecting people, so will the Internet of Things by interconnecting everyday things
A modest improvement of 1% would contribute significantly to the return on investment of the capital and operational expenses incurred by deploying the Industrial Internet.

- In aviation, the fuel savings of 1% per annum relates to saving $30 billion.

- Similarly, 1% fuel savings for the gas-fired generators in a power station returns operational savings of $66 billion.

- In the Oil and Gas industry, the reduction of 1% in capital spending on equipment per annum would return around $90 billion.

- The same holds true in the agriculture, transportation, and health care industries.
Key benefits adopters want from the Industrial Internet:
1. increased profits,
2. increased revenue flows, and
3. lower operational expenditures.

Some success stories:
- The success experienced by Thames Water, it uses the IIoT for remote asset management and predictive maintenance: anticipate equipment failures and respond quickly to any critical situation.
- Innovative projects from using drones and autonomous vehicles to inspect Oil and Gas lines in inhospitable areas.
- Logistics is the management of the flow of things between the point of origin and the point of consumption in order to meet requirements of customers by using embedded RFID tags.
Integration of horizontal cyber physical systems and vertical business processes

To align to Industry 4.0, companies must undergo digital transformation: digitalising, automating, and connecting all machines, manufacturing and business processes.
Industrial systems combine a mixture of sensors, actuators, logic and computing components, and networks to allow them to interconnect and function.

Industrial systems become Industrial Internet systems when they become connected to the Internet and integrate with enterprise systems, for improving business.
Initiatives around the world

- **Germany** – Industrie 4.0
- **USA** – Industrial Internet Consortium, Advanced Engineering Partnership
- **UK** – Catapults, UK Digital Strategy, Made Smarter review
- **France** – La Nouvelle France Industrielle
- **Japan** – Industrial Value Chain Initiative, New Robot Strategy, RRI
- **China** – Made in China 2025
- **Netherland** – Smart Industry
- **Belgium** – Made Difference
- **Spain** – Industrie Conectada
- **Italy** – Fabbrica Intelligence
- **Russia** – National Technology Initiative
- **Australia** – Industry 4.0 TestLabs

*Bring back manufacturing from “low-wage” countries back to “high-wage countries***
Initiatives around the world

Bring back manufacturing from “low-wage” countries back to “high-wage countries”

Food for thought: what would be the best strategy for decision makers of “low-wage” countries?

Profound implication
What Is the Industrial Internet?

The Industrial Internet provides:

- a way to get better visibility and insight into the company’s operations and assets through integration of advanced technologies.
- a method of transforming business operational processes by using as feedback the results gained from interrogating large data sets through advanced analytics.

The business gains are achieved through:

- operational efficiency gains and accelerated productivity,
- results in reduced unplanned downtime and optimized efficiency, and profits.
Why Industry 4.0 now?

The complexity of industrial systems has outpaced the human operator’s ability

Enabling technologies are maturing and widely available.
**Industrial Internet** concentrates on manufacturing, transportation, public sector, and related industrial systems such as:

- Agriculture, Aviation, Energy production, Health care,
- Manufacturing, Logistics, Transportation, and others

**Industry 4.0 or I4.0** is the fourth industrial revolution: the Industrial Internet with focusing more on the manufacturing environment.

- GE (General Electric) coined the name “Industrial Internet” as their Industrial IoTs (IIoTs)
- Cisco termed it the Internet of Everything (IoE)
- Others called it Internet 4.0 or other variants.

Usage: Industrial Internet of Things (IIoT) = *Industrial* Internet = Industry 4.0 = I4.0
Moving on to
Industry 4.0 manufacturing
Scenario - A smart machine fills each bottle with the same base ingredients. Each variant of the brand may have different colour additives of perfume as required for an intended market.

In traditional manufacturing: a production line is required for each individual product: A dispenser machine would fill the bottle with the required mixture of ingredients as required.
Design a production line to produce all products even though they differ in label, colour, and perfume.

Products on the production line are identified together with their status, their history, and what stage of production they must next pass through.

The machine must be able to identify each product traversing the production line and what to do with the product.

Each product has its own ID and relevant data stored in RFID tags.

Alasdair Gilchrist, Industry 4.0: The Industrial Internet of Things, Apress 2016
An intelligent, self-regulating, automated manufacturing process on cyber-physical production systems,

It can produce one or more products simultaneously.

Products navigate themselves through the product lifecycle via the cyber-physical production systems CPS without direct human intervention.

- Decentralised information
- Self-organisation
- Communication and cooperation among modules
Industrial Internet – underpin technologies

- Advanced Sensing Technology: Wireless and Sensor Miniaturization
- Ubiquitous Networking Technologies: Software Defined networking and Network Function Virtualization
- Intelligent Manufacturing Technologies: Platform, System, Product Life Cycle
- Digital Twins and Simulation
- Fog/Edge and Cloud Computing
- Big Data and Analytics
- Robots
- Additive Manufacturing (3D printing)
Sensor Technology innovation

- **Miniaturization of sensors and components**
  - Sensors can be reduced to the size of a grain of sand, can now be embedded anywhere and in anything: our bodies, our clothes, food packages.

- **Widespread use of multi-sensor systems**
  - In smartphones, systems-on-a-board, and even systems-on-a-chip (SoC): Apple iPhone, Raspberry Pi, Arduino for sensing and influence their environment.

- **Availability of autonomous and mobile sensors**

*Intelligent devices.* Any device with some intelligent capability for gathering information we wish to harvest, for example, sensors, actuators, engines, machines, components, even the human body, etc.
A cyber-physical systems (CPS) is a system that integrates all three features: computation, networking, and physical processes.

- Integrate with their environment
- Embedded computers and networks monitor and control the physical processes via feedback loops.
- CPPSs exchange of information over the entire lifecycle of a product, as data transfers seamlessly from system to system.
- Machines monitor one another and make decentralised decisions about production and maintenance.

Examples: Human operators, Smart factories, Smart phones, Robots, Intelligent Manufacturing Lines
Digital twin

- Digital twin: a digital version of a physical thing

- Digital twin of a product or process: The digital twin refers to a digital model of a particular product or process that includes:
  - design specifications and
  - engineering models describing its geometry, materials, components, assembly and behaviour, and
  - the as-built and operational data unique to the specific physical asset which it represents.

Creating machine understandable virtual image of an entity -> entering the virtual world
Digital Twins

- Digital Twin of a Product
- Digital Twin of a Production Asset
- Digital Twin of a Process
- Digital Shadow of a Factory
- Digital Threat

A Factory
NIST: CLOUD computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.
**Implications:**
- Centralized control
- Programmability
- Embrace Network Function Virtualization
- Virtualization
- Autonomous devices management

Virtual Networks are created just like the way Virtual Machines are created

**Figure 1 - Software-Defined Networking – A high level architecture**

**FAIR, Hue, Vietnam 2019**
Network Function Virtualization (NFV)

NFV is the method and technology that enables one to replace physical network devices performing specific network functions with one or more software programs executing the same network functions while running on commodity hardware.
Big data refers to datasets whose size is beyond the ability of typical database software tools to capture, store manage, and analyse. This definition is intentionally subjective.

- One query search in Google results in over 1M of web pointers.
- Everyday, 2.5 quintillion (or Exa = $10^{18}$) bytes of data are created and 90% of the data in the world today were produced within the last two years.
- Flickr, a public picture sharing site, which received 3.6 million of pictures per day on average. This is equivalent to 7.2 terabytes storage every single day.

Big Data is characterized by
- Volume,
- Velocity
- Variety
Example of Architecture Pattern: 3 Tiers

The edge tier is where data from all the endpoints is collected, and transmitted over the proximity network to a border gateway.

The platform tier receives data from the edge tier over the access network. Responsible for data transformation and processing. Also for managing control data flowing in the other direction.

The enterprise tier implements the application and business logic for decision support. And end-user interfaces. Host most of the application and business functions.
Main points so far
WHERE DO WE GO FROM HERE?

Application and Services

Data Analytics
Machine Learning

Big Data and Data Management

Management
• Fault
• Configuration
• Performance
• Security
• Accounting

Adapting

Automation

Real-time

Gateway

Production line
Manufacturing Plan
Cyber Physical System

Data Analytics in Production Systems

Data Lifecycle = data gathered by the device ==> data store ==> analytic systems ==> data scientists ==> process ==> the device

FAIR, Hue, Vietnam 2019
### Huge data from the environment IoTs or networks

**NFV: Software Entities/Functions/Services**
- Virtual resources generated on-demand
- Deployed anytime, anywhere, anyplace
- Generate data
- Execute actions

**SDN**
Creating dynamic connectivity infrastructure
- Logically centralized control
- Programmability
- Automated configuration and management

### Cloud Computing
- Resources virtualization and provision
- Repository for big data
- Computing power for data processing and value extraction

### AI Machine Learning and Data Analytics

- for generating value services
Data Analytics - Questions

**Prescriptive**
- What is the best action?
- What if?

**Predictive**
- What will happen next?
- What is the pattern?

**Diagnostic**
- Where to look?
- Why did it happen?

**Descriptive**
- When, where?
- What happened?
Data Analytics is the science of examining raw data with the purpose of drawing conclusions about that information (uncovering patterns, correlations, and other features).

Data science (combination of statistics, mathematics, programming, problem-solving, and data management) is more about exploring unknowns, while data analytics emphasizes on discovering answers to questions being asked.
To solve a problem on a computer, we need an algorithm. An algorithm is a sequence of instructions that should be carried out to transform the input to output. For example, one can devise an algorithm for sorting.

For some tasks, however, we do not have an algorithm. E.g., telling spam emails from legitimate ones. We know what the input is an email document. We know what the output should be: a yes/no output. But we do not know how to transform the input to the output.
We do not have an algorithm but have lots of data, what we want is to "learn" what constitutes spam from the data.

Machine learning is to extract automatically the algorithm for the task that we do not have an algorithm.

Machine learning algorithms build a mathematical model based on sample data to make predictions or decisions without being explicitly programmed to perform the task.
Machine Learning and Data Analytics Process

- Process Industry
  - Datasets preparation
    - Sample and Variable selection
  - Data Pre-processing
    - Model Training
    - Model Validation
  - ML Algorithms
    - Data Analytics
      - Classification
      - Regression
      - Clustering
      - Feature Extraction
      - Abnormal Detection
      - Predictive Analysis
      - Visualization
      - Process Monitoring
- Real-time data
  - Knowledge discovery
  - Knowledge Automation
  - Process Improvement
  - Decision support

Adapted from Z. Ge et al.: Data Mining and Analytics in the Process Industry: The Role of Machine Learning, IEEE Access, 2017
Machine Learning. A specific subset of AI that trains a machine how to learn, makes it possible to quickly and automatically produce models that can analyze bigger, more complex data and deliver faster, more accurate results.
Learning in Machine Learning: Supervised Learning

Sample data

\( x_k \)

Selected model. Parameters
\( \omega_1, \omega_2, \omega_n \)

Adjust parameters

Train to minimise \( \varepsilon \) or until correct classification

\( y_k, y'_k, \varepsilon \)
Learning in Machine Learning: Unsupervised Learning

Sample data $x_k$ → Selected model. Parameters $\omega_1, \omega_2, \omega_n$ → Learning rule

Train based on sample Similarity or difference by clustering or discriminating
Widely Used Machine Learning in Process Industry

Supervised Learning
- Partial Least Squares
- Artificial Neural Networks
- Support Vector Machine
- K-nearest Neighbour
- Decision Tree

Unsupervised Learning
- Principal Component Analysis
- K-means
- Kernel Density Estimate
- Self Organizing Map
- Support Vector Data Description

Reinforcement Learning
- Performance Index
- Prediction
- Fault Classification and Identification
- Process Monitoring Fault diagnosis

Semi-supervised Learning

Transfer Learning

Deep Learning
is of particular interest

## Frequently used ML algorithms

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-Nearest Neighbors</td>
<td>Classification</td>
</tr>
<tr>
<td>Naive Bayes</td>
<td></td>
</tr>
<tr>
<td>Support Vector Machine</td>
<td>S</td>
</tr>
<tr>
<td>Classification</td>
<td>S</td>
</tr>
<tr>
<td>Regressions</td>
<td></td>
</tr>
<tr>
<td>Linear Regression</td>
<td>Regression</td>
</tr>
<tr>
<td>Classification and Regression Trees</td>
<td>S*</td>
</tr>
<tr>
<td>Random Forests</td>
<td>S*</td>
</tr>
<tr>
<td>Bagging</td>
<td>S*</td>
</tr>
<tr>
<td>K-Means</td>
<td>Clustering</td>
</tr>
<tr>
<td>Density-Based Spatial Clustering 0f</td>
<td>U*</td>
</tr>
<tr>
<td>Principal Component Analysis</td>
<td>Applications with Noise</td>
</tr>
<tr>
<td>Canonical Correlation Analysis</td>
<td>Feature extraction</td>
</tr>
<tr>
<td>Feed Forward Neural Network</td>
<td>Regression/Classification/</td>
</tr>
<tr>
<td>One-class Support Vector Machines</td>
<td>Clustering/Feature extraction</td>
</tr>
<tr>
<td></td>
<td>Anomaly detection</td>
</tr>
</tbody>
</table>

Adapted from Z. Ge et al.: Data Mining and Analytics in the Process Industry: The Role of Machine Learning, IEEE Access, 2017
A road block: Cybersecurity, Safety, Security and Privacy
1. Network

2. Infrastructure

SDN issues: controller security issues, flooding flow switches, insecure interfaces (North, South, East, West).
Secure billion of IoT devices.

3. Reality

Security holes everywhere: there are holes everywhere (wherever we turn our attention there will be holes to be plugged.)
Expertise on specific concern per network layer experts, per specific sub-syst, experts, per application experts, per security issue experts, etc.
Lack global view and assurance. No overall model that can give assurance. Sum of assurance of component concerns do not provide assurance of the overall system.
No information sharing or bridging across concerns to build models and standards.

Cybersecurity and Risk Assessment

Software-defined Security

Centralized Control/Policy
Programmability
Automation
Virtualization

Software Defined Security Architecture and Implementation

SD Security

Security Controller

SDN Orchestration
Compute Orchestration
Storage Orchestration
NFV Orchestration
A risk is in essence the product of threats, vulnerabilities, and the consequences of the exploitation of vulnerabilities by the threats (i.e., the impact of threat)

Three questions are answered during a quantitative risk analysis:
- A scenario $s_i$, (i.e., what can go wrong?)
- The probability $p_i$, of $s_i$, (the probability that the scenario is realized)
- The consequence of $x_i$ of $s_i$.

But that is not enough! When an incident occurs we want to know where it hurts and by how much.
• Challenges and Opportunities
ML Intelligence
Cyber Physical System
Cybersecurity
Direction and Action
ML Challenges

How to put intelligence into a ML algorithm?

What are the challenges in all ML algorithms?

- Is there enough data to learn what we expect or desire?
- Given the data, does the chosen model have adequate dimensions or parameters to discover all that is contained in the data?
- Which ML algorithm is most suitable for a given set of big data?
- Which learning method is suitable for a particular task relative to the selected data?
ML has become just a tool, more can be done on it
Challenges in Data Intelligence

- What value do we want to extract?
- What value we would like to extract from the data?
- What data should be collected

Assessment Strategy

We can not detect unknown (if we do not know what we look for)

So how do we detect unknown features?

Self-organization, self-clustering methods? BUT HOW?

New projection from existing data – How about a new dimension?

Introducing random events to find unknowns
On Cyber Physical Systems

Automation & Real-time

Security implication on all products and services
Bring back manufacturing from “low-wage” countries back to “high-wage countries?”
What types of human resources are needed for I4.0? And how do we prepare for the need?

1. Data Managers: managing data and data framework
2. Security experts: protection of assets
3. Data Scientists who understand business to extract value from data
4. IT experts (network, database, application): integrated knowledge for overall system
5. Business analysis – creating businesses
Question?

What would be the best strategy for decision makers of “low-wage” countries?
• **I4.0 is here to stay** - Best option: Implement and deploy best I4.0 to support the economy

• **Data Analytics is absolutely essential**: Need top researchers/experts to extract value from data for business

• **Building blocks**: Invest in building smart factories, supply chains, and logistics

• **Security and Privacy**: Built in for the whole lifecycle of any product or service

• **Invest in people wisely for the future** - Education
Reference
A. Gilchrist, Industry 4.0: The Industrial Internet of Things, Apress 2016

Thank You