GOI ESKOLA POLITEKNIKOA
FACULTY OF ENGINEERING

LEVERAGING MULTIVARIATE ANALYSIS TO DETECT ANOMALIES IN INDUSTRIAL CONTROL SYSTEMS

DAT300 presentation

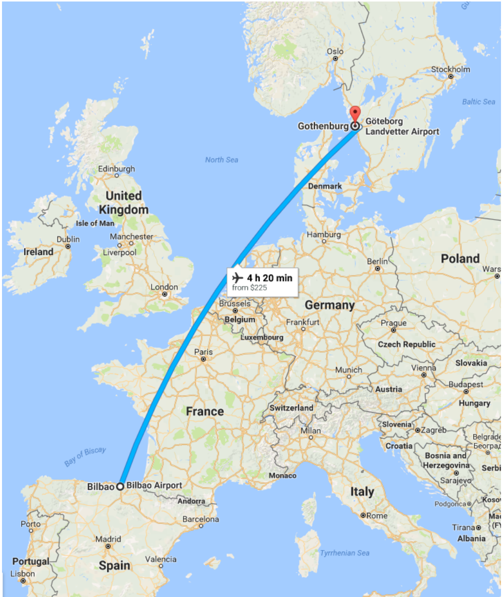
Mikel Iturbe

Electronics and Computing Department
Faculty of Engineering – Mondragon University

Chalmers University of Technology, Gothenburg, Sweden
September 15, 2016

<Prologue>

About me


Oslo
Stockholm
Baltic Sea
Gothenburg
Göteborg
Landvetter Airport
North Sea
Edinburgh
Denmark
United Kingdom
Isle of Man
Dublin
Ireland
Manchester
Liverpool
Hamburg
Berlin
Poland
4 h 20 min
from $225
London
Brussels
Belgium
Germany
Frankfurt
Prague
Luxembourg
Czech Republic
Paris
Slovakia
Munich
Vienna
Budapest
Austria
Hungary
France
Switzerland
Slovenia
Zagreb
Milan
Croatia
Bay of Biscay
Bosnia and Herzegovina
Sarajevo
Monaco
Bilbao
Bilbao Airport
Italy
Andorra
Rome
Porto
Barcelona
Madrid
Tirana
Albania
Tyrrhenian Sea
Portugal
Spain
Valencia
Lisbon

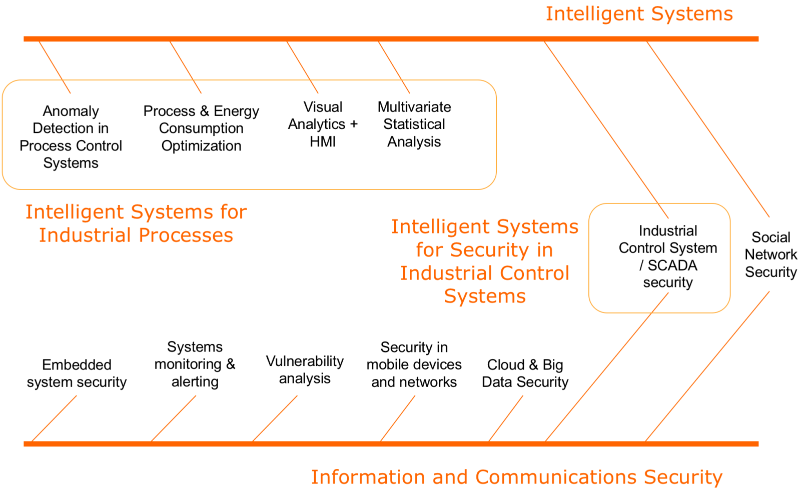
About me

- Born in Bilbao, Basque Country, 1987
- BSc in Computer Engineering (Mondragon Unibertsitatea 2008-2012)
- MSc in ICT Security (UOC, UAB, URV 2012-2013)
- PhD in ICS Security (Mondragon Unibertsitatea 2013-2017?)

About us: Mondragon Unibertsitatea

- Small, private, non-profit university in the Basque Country
- Founded in 1997 (1943)
- Some data (14/15)
    - 4 faculties
    - 3513 undergrad students
    - 615 Master students
    - 112 PhD students
- Cooperative university
- Transfer oriented research

About us: Telematics team at Mondragon Unibertsitatea

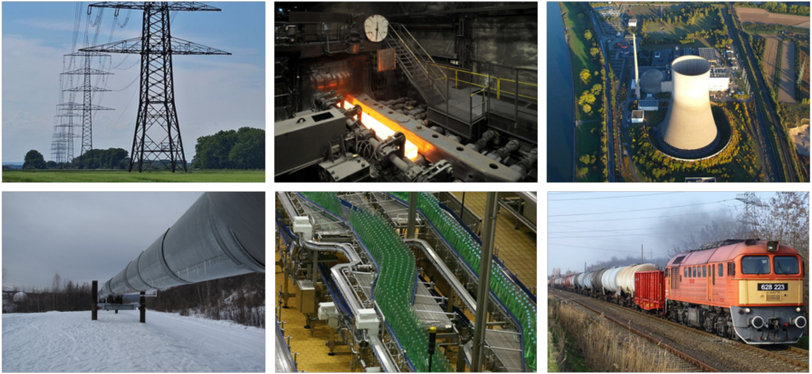


</Prologue>

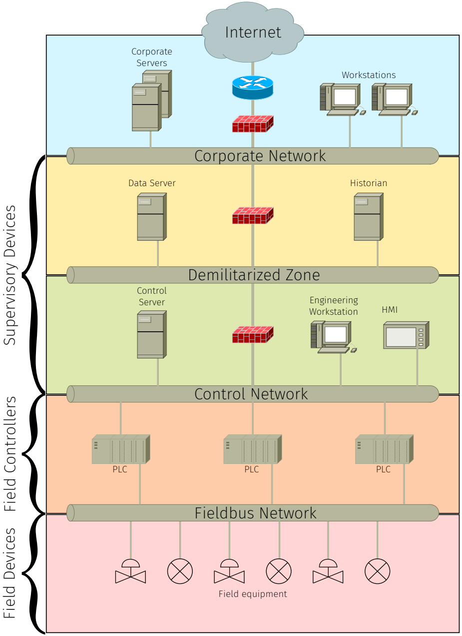
Agenda

Introduction

Industrial control systems



CC-BY-SA 3.0 Kreuzschnabel, Schmimi1848, Wolkenkratzer, Brian Cantoni, Hermann Luyken, Beroesz


Internet
Corporate Servers
Workstations
Corporate Network
Data Server
Historian
Demilitarized Zone
Control Server
Engineering Workstation
HMI
Control Network
PLC
PLC
PLC
Fieldbus Network
Field equipment
Supervisory Devices
Field Controllers
Field Devices

ICS vs. IT

	ICS networks	IT networks
Primary function	Control of physical equipment	Data processing and transfer
Applicable Domain	Manufacturing, processing and utility distribution	Corporate and home environments
Hierarchy	Deep, functionally separated hierarchies with many protocols and physical standards	Shallow, integrated hierarchies with uniform protocol and physical standard utilisation
Failure Severity	High	Low
Reliability Required	High	Moderate
Round Trip Times	250µs–10 ms	50+ ms
Determinism	High	Low
Data Composition	Small packets of periodic and aperiodic traffic	Large, aperiodic packets
Temporal consistency	Required	Not Required
Operating environment	Hostile conditions, often featuring high levels of dust, heat and vibration	Clean environments, often specifically intended for sensitive equipment
System lifetime (years)	Some tens	Some
Average node complexity	low (simple devices, sensors, actuators)	high (large servers/file systems/databases)

Brendan Galloway and Gerhard Hancke. Introduction to Industrial Control Networks. *IEEE Communications Surveys & Tutorials*, 15(2):860–880, 2012
Manuel Cheminod, Luca Durante, and Adriano Valenzano. Review of Security Issues in Industrial Networks. *IEEE Transactions on Industrial Informatics*, 9(1):277–293, 2013

Anomaly Detection Systems

Intrusion Detection System

- Mechanisms that monitor network and/or system activities to detect suspicious events that occur in them.
- Main classification criteria in IDSs
  1. Detection mechanism
     - Signature-based
     - Anomaly Detection Systems (ADS)
  2. Scope
     - Host
     - Network
  3. ICS Scope
     - Network
     - Process

ADSs on ICSs

- Active research topic
- Data-driven methods gaining traction

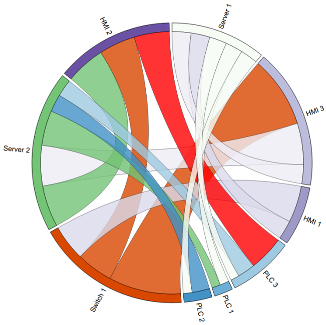
Bonnie Zhu and Shankar Sastry. SCADA-specific intrusion detection/prevention systems: a survey and taxonomy. In *Proceedings of the 1st Workshop on Secure Control Systems (SCS)*, 2010

Robert Mitchell and Ingray Chen. A Survey of Intrusion Detection Techniques for Cyber Physical Systems. *ACM Computing Surveys*, 46(4), April 2014

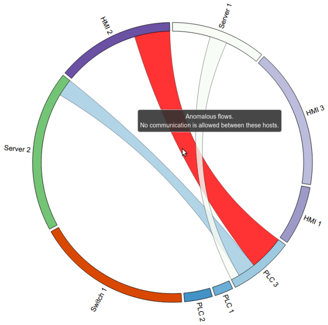
Gaps in literature

- We found a couple of relevant gaps in the literature.
    1. Lack of visualizations
    2. Almost no network & process level ADSs

Visual Network Flow Monitoring



(a) Forbidden flow between PLC 1 and HMI 2.



(b) Detail of the forbidden flow.

Mikel Iturbe, Iñaki Garitano, Urko Zurutuza, and Roberto Uribeetxeberria. Visualizing Network Flows and Related Anomalies in Industrial Networks using Chord Diagrams and Whitelisting. In *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, volume 2, pages 99–106, Feb. 2016

Visual Network Flow Monitoring

Able to detect, based on whitelisting:

- Forbidden connection
- Forbidden port
- Incorrect flow size (DoS)
- Missing host

Gaps in literature

- We found a couple of relevant gaps in the literature.
  1. ~~Lack of visualizations~~
  2. Almost no network & process level ADSs

Gaps in literature

"In order to make IDSs effective in protecting this kind of systems, it is then needed a set of multilayer aggregation features to correlate events generated from different sources (e.g. correlating events coming from the process network of a remote transmission substation with events coming from the office network of a control center) in order to detect large scale complex attacks. This probably represents the next research challenge in this field."

Ettore Bompard, Paolo Cuccia, Marcelo Masera, and Igor Nai Fovino. Cyber vulnerability in power systems operation and control. In *Critical Infrastructure Protection*, pages 197–234. Springer, 2012

Multivariate Statistical Process Control

Multivariate data

	V_1	V_2	V_3	...	V_m
O_1					
O_2					
O_3					
$\vdots$					
$\vdots$					
$\vdots$					
O_n					

- ICSs are multivariate by nature.

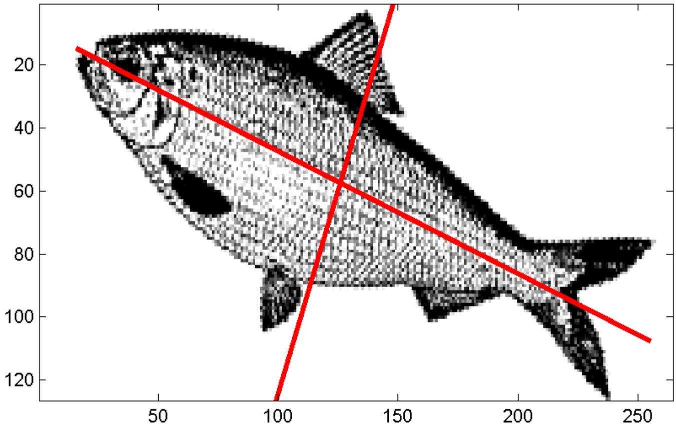
Multivariate data

It is not easy to monitor...

- If variables are in their normal operation constraints
- Correlations between different variables

But, information can be expressed in a (smaller) set of non-measurable variables called Latent Variables or Principal Components

Principal Component Analysis (PCA)

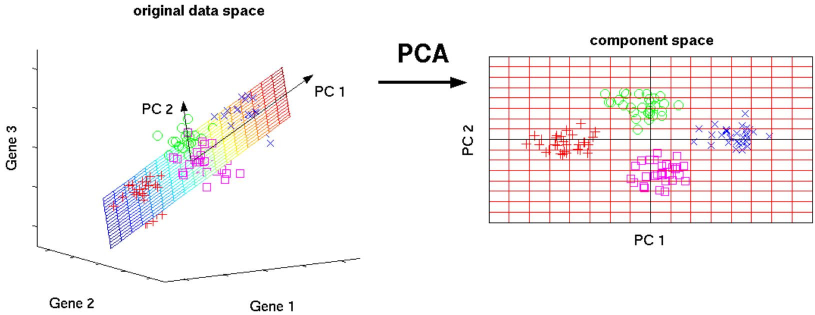


CC-BY-SA 2.5 Charles Albert-Lehalle

Principal Component Analysis (PCA)

- Dimensionality Reduction Algorithm
- Linear combination of variables
- Maximizes variance

Principal Component Analysis (PCA)



CC-BY 2.0 Matthias Scholz, Approaches to analyse and interpret biological profile data. PhD Thesis. University of Potsdam, 2006

Principal Component Analysis (PCA)

$$X = T_A P_A^t + E_A$$

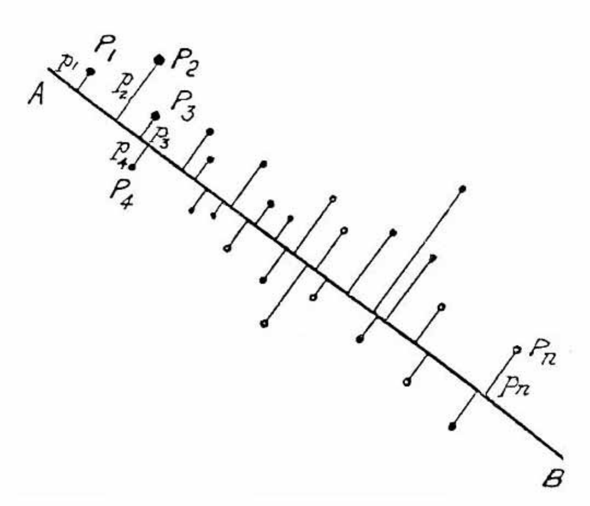
$$\begin{bmatrix} O_{11} & \dots & O_{1m} \\ O_{21} & \dots & O_{2m} \\ O_{31} & \dots & O_{3m} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ O_{n1} & \dots & O_{nm} \end{bmatrix} = \begin{bmatrix} O'_{11} & O'_{12} \\ O'_{21} & O'_{22} \\ O'_{31} & O'_{32} \\ \vdots & \vdots \\ \vdots & \vdots \\ O'_{n1} & O'_{n2} \end{bmatrix} \begin{bmatrix} V'_{11} & \dots & V'_{1m} \\ V'_{21} & \dots & V'_{2m} \end{bmatrix} + \begin{bmatrix} e_{11} & \dots & e_{1m} \\ e_{21} & \dots & e_{2m} \\ e_{31} & \dots & e_{3m} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ e_{n1} & \dots & e_{nm} \end{bmatrix}$$

Principal Component Analysis (PCA)

$$X = T_A P_A^t + E_A$$

$$\underset{\text{Original values}}{\begin{bmatrix} O_{11} & \dots & O_{1m} \\ O_{21} & \dots & O_{2m} \\ O_{31} & \dots & O_{3m} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ O_{n1} & \dots & O_{nm} \end{bmatrix}} = \underset{\text{Scores}}{\begin{bmatrix} O'_{11} & O'_{12} \\ O'_{21} & O'_{22} \\ O'_{31} & O'_{32} \\ \vdots & \vdots \\ \vdots & \vdots \\ O'_{n1} & O'_{n2} \end{bmatrix}} \underset{\text{Loadings}}{\begin{bmatrix} V'_{11} & \dots & V'_{1m} \\ V'_{21} & \dots & V'_{2m} \end{bmatrix}} + \underset{\text{Residuals}}{\begin{bmatrix} e_{11} & \dots & e_{1m} \\ e_{21} & \dots & e_{2m} \\ e_{31} & \dots & e_{3m} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ e_{n1} & \dots & e_{nm} \end{bmatrix}}$$

PCA: Residuals



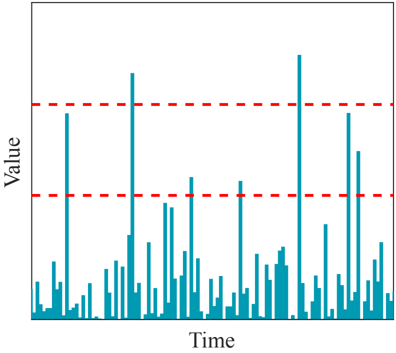
PD - Pearson, K. 1901. On lines and planes of closest fit to systems of points in space. Philosophical Magazine 2:559-572.

Multivariate Statistical Process Control

- Statistical Control
- Process-agnostic
- We monitor the scores and the residuals on control charts.
- Two univariate statistics:

$$D_n = \sum_{a=1}^{A} \left(\frac{t_{an} - \mu_{\mathbf{t}_a}}{\sigma_{\mathbf{t}_a}} \right)^2 \; ; \; Q_n = \sum_{a=1}^{A} (e_{nm})^2$$

Control Charts

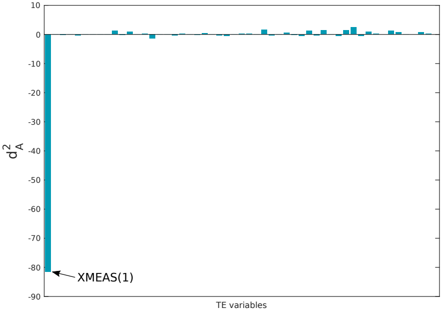


Anomaly Detection

- Monitoring of control charts
- If three consecutive observations go of bounds, the event is flagged as anomalous

Anomaly Diagnosis

- Once an anomaly is flagged, we diagnose its cause
  - Contribution (oMEDA) plots



José Camacho. Observation-based missing data methods for exploratory data analysis to unveil the connection between observations and variables in latent subspace models. *Journal of Chemometrics*, 25(11):592–600, 2011

Application to Network Anomaly Detection

- MSPC-based techniques can be used for network anomaly detection.
- Variable parametrization.
    - Logs

José Camacho, Gabriel Maciá Fernández, Jesús Díaz Verdejo, and Pedro García Teodoro. Tackling the Big Data 4 vs for anomaly detection. In *Computer Communications Workshops (INFOCOM WKSHPS), 2014 IEEE Conference on*, pages 500–505, April 2014. doi: 10.1109/INFCOMW.2014.6849282

José Camacho, Alejandro Pérez Villegas, Pedro García Teodoro, and Gabriel Maciá Fernández. PCA-based multivariate statistical network monitoring for anomaly detection. *Computers & Security*, 59:118–137, 2016. ISSN 0167-4048. doi: http://dx.doi.org/10.1016/j.cose.2016.02.008

And in ICSs?

- When looking at Process Data, we might be able to distinguish intrusions from disturbances using MSPC

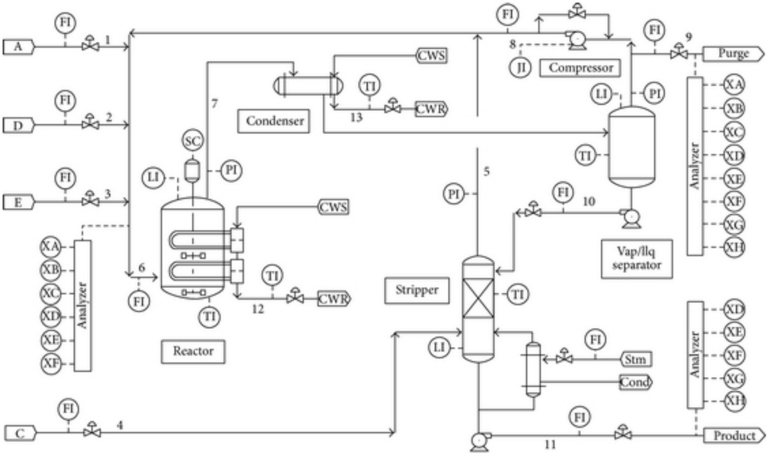
Mikel Iturbe, José Camacho, Iñaki Garitano, Urko Zurutuza, and Roberto Uribeetxeberria. On the feasibility of distinguishing between process disturbances and intrusions in process control systems using multivariate statistical process control. In *2016 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN 2016)*, 2016

Work Hypothesis

Therefore, it seems natural to link both worlds, and create a unified ADS for ICSs.

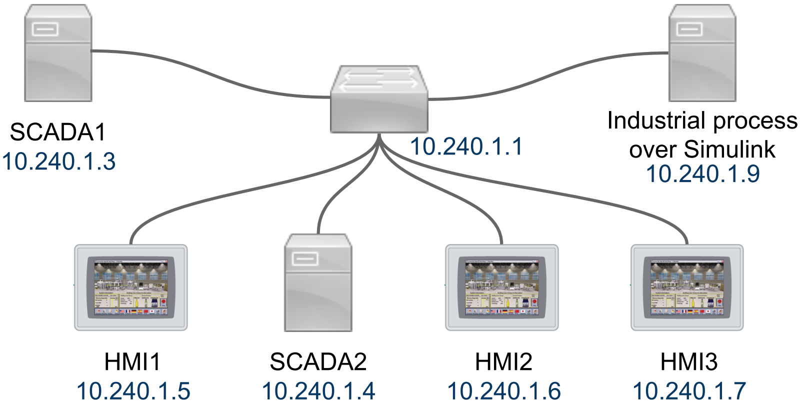
Ongoing work

Process: Tennessee-Eastman



James J Downs and Ernest F Vogel. A plant-wide industrial process control problem. *Computers & Chemical Engineering*, 17(3):245–255, 1993

NETWORK


SCADA1
10.240.1.3
10.240.1.1
Industrial process
over Simulink
10.240.1.9
HMI1
10.240.1.5
SCADA2
10.240.1.4
HMI2
10.240.1.6
HMI3
10.240.1.7

Challenges

- Timestamp synchronization
- Data processing complexity

Conclusions

Conclusions

- Anomaly Detection in ICSs is an active research field
- Security visualizations in the field are still in their infancy
- Multivariate Analysis can help finding process-level anomalies
- Network variable parametrization opens the way to a multi-level, process-agnostic, ADS for ICSs.

THANK YOU.

miturbe@mondragon.edu

iturbe.info