Design and implementation of secure industrial network

HALLER Piroska

"Petru Maior" University
Tirgu Mures, Romania
e-mail: phaller@upm.ro

6th Int. Conf. on Mathematics and Informatics,

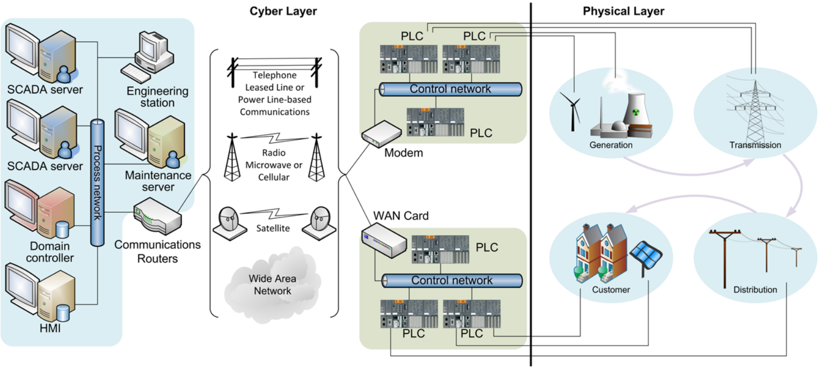
September 7-9, 2017

Open questions

- **Why is Industrial Networking Different?**
- **Control theory or graph theory?**
- **Network or physical process monitoring?**
- **Can increase the detection accuracy by design?**

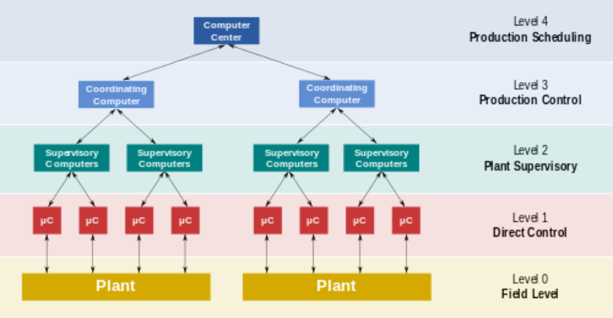
Industrial Control Networks

- Architecture includes the cyber and physical domains
  - The physical process: chemical plant, electricity grid, ...
  - Programmable Logical Controllers (PLC)
  - Master Terminal Units (MTU - SCADA servers)
  - Human Machine Interfaces (HMI)
  - Communication infrastructure



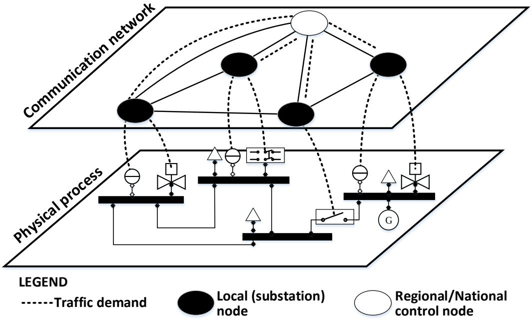
The Evolution of Industrial Networks

- Combine the isolated control system and public networks
- Several distributed systems running in parallel
- Allow real time remote access by a wide variety of users
- Optimal control in the interconnected system
- Provide real-time diagnosis, self-healing
- Change of the system parameters, to assures a balanced load



ICN vs traditional Computer Networks

- ICN are connected to physical equipment: failure of industrial networks can have severe repercussions
- ICN have strong determinism (transmission and reply are predictable)
- ICN communicating edge nodes are well-known
- ICN include strict real-time requirements
- ICN have longer lifetimes (at least ten years)



IT Network security

- Multilevel security
- Security zones, Demilitarized zones
- Secure communication channels
- Firewall and packet filter
- Intrusion detection, Intrusion protection (IDS, IPS)
- Logging and audit
- Risk analyses and management

The network level IDS, IPS has no access to the state variables of the physical process

Intrusion Detection Systems at the Process Level

- **Control theory based** - Centralized
  - Assume that dynamical system model is known
  - Synchronous, and secure sampling of the whole system state
  - Solve the model in real time and compare the estimated and measured values
- **Time series analysis** - Distributed
  - Use sensitivity analysis to identify sensitive variables to specific interventions
  - Adopts the cross-association assessment to group the process variables
  - Optimal design of the process level IDS

Sensitivity analysis

- Builds on system dynamics
- Incorporates process behavior (time-based dynamics)
- Records measurements of observable variables in the absence of an intervention, and in the presence of a specific intervention
- Builds a map of cyber attack impact propagation
- Assesses the impact of cyber attacks on heterogeneous ICS

Cyber attack models

- **Replay attack:** the attacker replays recorded packets from the past.
- **False data injection:** the attacker injects modified data packets.
- **Denial of Service (DoS):** the attacker is aimed to disrupt a service.

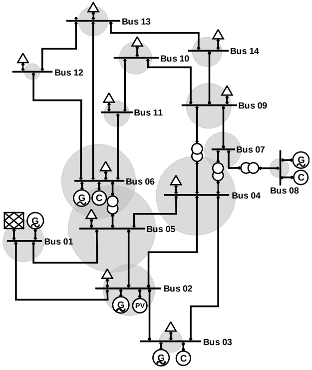
Sensitivity analysis

- Y_j^0, be a vector containing w measurements for observed variable j without intervention, $J = \{1, 2, ..., j, ...m\}$
- Y_j^i denote the recorded vector for an intervention i on the j-th observed variable
- $i = (var, type, parameters)$, the intervention tuple i as a specific intervention type applied to a selected variable, $I = \{1, 2, ..., i, ...n\}$
- c_{ji}, sensitivity index

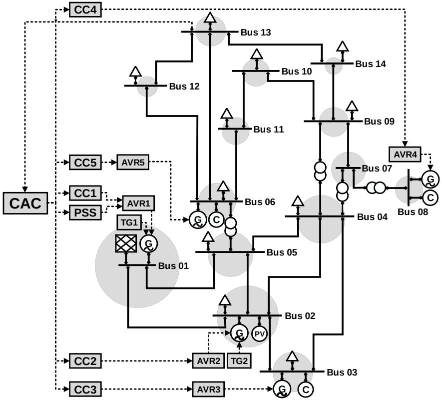
$$c_{ji} = \frac{std(Y_j^i)}{std(Y_j^0)}, \forall j \in J, \forall i \in I$$

- $\boldsymbol{C}$, represents the sensitivity index matrix

Sensitivity analysis using CAIA [1]



Without control loop



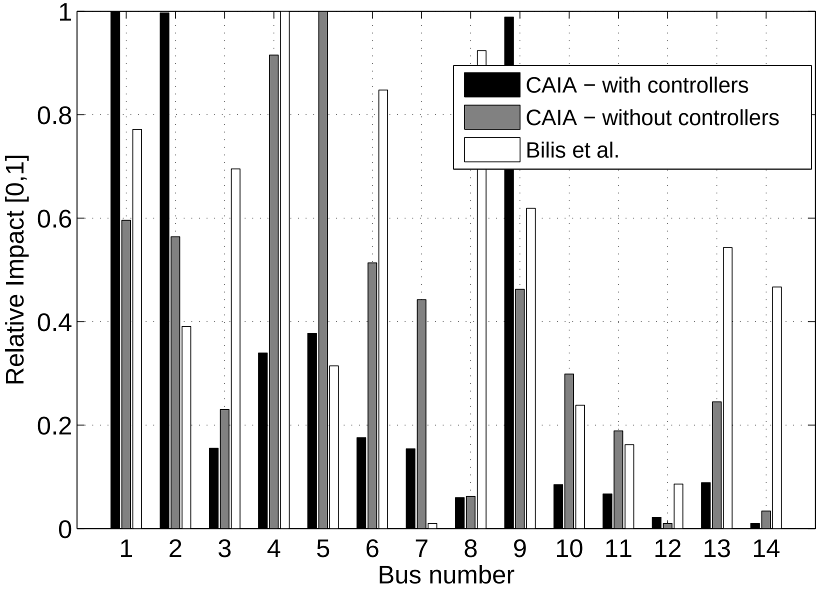
With control loop

[1]Genge B., Kiss I., Haller P.: A system dynamics approach for assessing the impact of cyber attacks on critical infrastructures

Sensitivity analysis based on graph metrics

- **Degree Centrality:** $C_D(v) = deg(v)/(n-1)$.
  Interpreted as number of vertices and edges that are directly influenced by the status of node v.
- **Eccentricity:** $C_E(v) = max[d(v,y)] \forall y \in V$, where $d(v,y)$ is the length of the shortest path connecting vertices v and y.
  Low eccentricity of node v suggests that all other nodes are in proximity.
- **Centroid Centrality:** $C_C(v) = d(v) - min[d(y)]$, where $d(v) = \sum d(v,y) \forall y \in V$ is the sum of distance to all others.
  A specific node has a central position within a graph region compare to any other node y, with a high density of interacting nodes.
- **Betweenness Centrality:** relative number of shortest path between any two vertices that pass through vertex v.

Compare the two sensitivity analysis



Selection of the Monitored Variables

- Identify the group of interventions that have the same impact on a group of observed variables
- Reduce the number of variables that need to be monitored
- Select the primary and secondary (resilient) monitoring variable set
- Define the appropriate intrusion detection method, based on the determined attack group
- Separate the random noise effect in the sensitivity index matrix from the intervention effect
- Obtain the binary version $\boldsymbol{B}$ of $\boldsymbol{C}$ using Expectation Maximization Clustering algorithm[2]

[2] McLachlan, G., and D. Peel. Finite Mixture Models

EM Algorithm for Gaussian Mixtures

- Create the unidimensional vector of the sensitivity values $D = \{x_1, x_2, ..x_N\}$
- Assumes that $p(x)$ is a finite mixture model with K components and z_k binary indicator variables

$$p(x|\Theta) = \sum_{k=1}^{K} \alpha_k p_k(x|z_k, \theta_k)$$

- $\alpha_k$are the mixture weights, $\sum_{k=1}^{K} \alpha_k = 1$
- Assumes that each of the K components a Gaussian density with parameters μ_k, σ_k
- Compute the membership weight of data point x_i in cluster k, given parameters $\Theta = \{\alpha_1, ..., \alpha_K, \theta_1, ..., \theta_K\}$

$$w_{ik} = p(z_{ik} = 1|x_i, \Theta) = \frac{\alpha_k p_k(x_i|z_k, \theta_k)}{\sum_{m=1}^{K} \alpha_m p_m(x_i|z_m, \theta_m)}$$

EM Algorithm for Gaussian Mixtures

- Iterative algorithm that starts from some initial estimate of Θ
- **Expectation step:** With the current Θ compute w_{ik} $\forall i, \forall k$
- **Maximization step:** Want to find the maximum likelihood estimate for parameter μ
- Use the membership weight to calculate new parameters:

$$N_k = \sum_{i=1}^{N} w_{ik}, \quad \alpha_k^{new} = \frac{N_k}{N}, \forall k$$

$$\mu_k = (\frac{1}{N_k}) \sum_{i=1}^{N} w_{ik} x_i, \forall k$$

- Stops when log-likelihood is not changing significantly

$$\sum_{i=1}^{N} log(p(x_i|\Theta)) = \sum_{i=1}^{N} (log(\sum_{k=1}^{K} \alpha_k p_k(x_i|x_k, \theta_k))$$

Cross-Associations

- Use a fully automatic cross-association simultaneously cluster the $\boldsymbol{B}$ matrix into disjoint row-column homogeneous groups
- Encode the binary matrix using the Minimum Description Length principle
- The costs are based on the number of bits required to transmit both the "summary" of the structure, as well as each rectangular region
- Determines the optimal number of row groups k and column groups l to obtain a uniform structure of cross-associate sub-matrices $B_{i,j}$ that minimize the cost[3]
- Use lossless code to describe a binary matrix - arithmetic coding

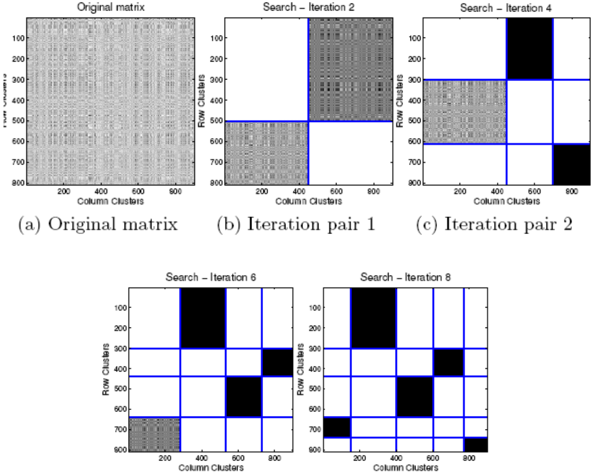
[3] Chakrabart D., Papadimitriou S., Modha D., and Faloutsos C., Fully automatic cross-associations

Cross-Associations

- Let an $n = axb$ binary matrix **B** with $n_1(B)$ number of nonzero entries and $n_0(B)$ number of zero entries
- The distribution of the elements $P_B(i) = n_i(B)/n(B), i = 0, 1$
- The total number of bits to encode the matrix will be:

$$CI(B) = \sum_{i=0}^{1} n_i(B) log \frac{n(B)}{n_i(B)}.$$

- The description of the whole matrix contains:
  - number of groups (k, l)
  - for each group the number of elements (a_i, b_j)
  - for each binary sub-matrix the code using $CI(B_{i,j})$ bits
- Iteratively search the optimal k^*, l^*, that minimize the total description length

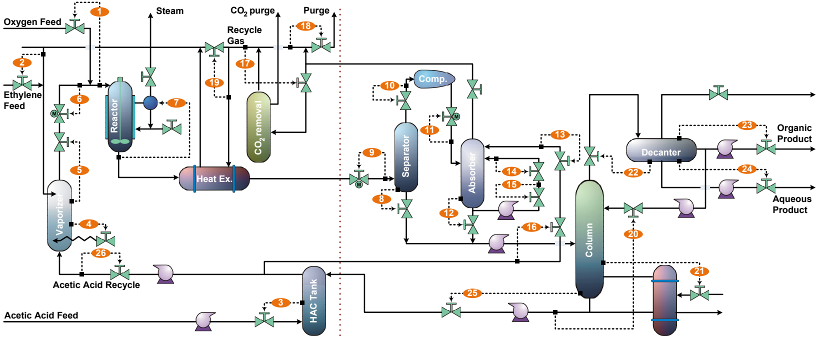


(a) Original matrix (b) Iteration pair 1 (c) Iteration pair 2

(d) Iteration pair 3 (e) Iteration pair 4

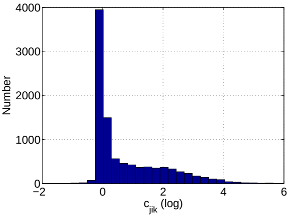
[4]Chakrabart D., Papadimitriou S., Modha D., and Faloutsos C., Fully automatic cross-associations

Chemical process

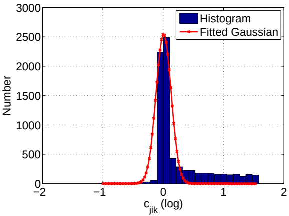


1040 attack experiments with different type and parameters

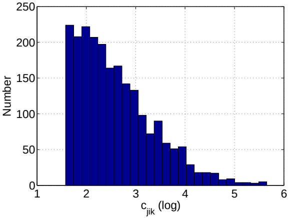
Selection of the Monitored Variables



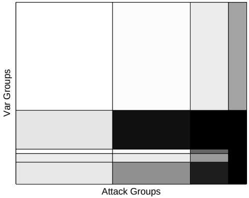
Complete histogram.



Non-sensitive components.

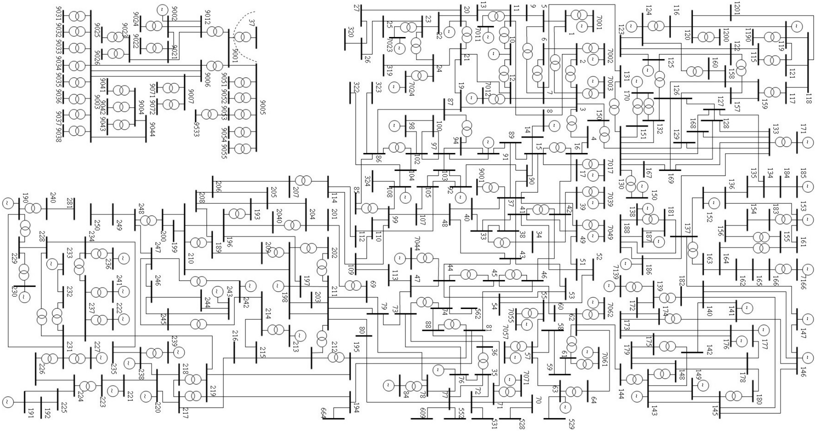


Sensitive components.

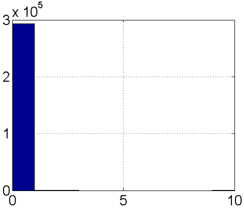


Cross-association.

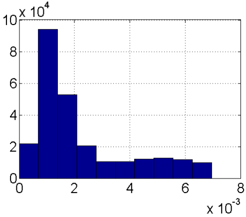
Electrical grid



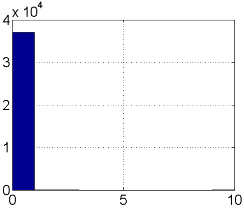
Selection of the Monitored Variables



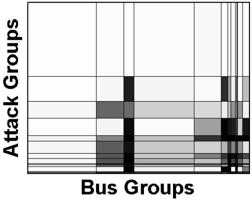
Complete histogram.



Non-sensitive components.



Sensitive components.



Cross-association.

IDS Network Design Problem

- Designing a resilient Industrial Network infrastructure,
- IDS engines are spread across the infrastructure to ensure the resilient monitoring
- Minimizes costs, while optimally selecting the shortest communication paths and the location of IDS devices
- Use different IDS device class for different intervention group (class)
- Provides K distinct back-up paths for each communication flow

Parameters

- I, J, D Flows (J), Routing Nodes (J), IDS device classes (D)
- c_{jl}^{L} The cost of bandwidth on link (j, l)
- c_{jv}^{V} The cost of detection bandwidth for v at RN j
- c_{i}^{P} Penalty cost for not monitoring flow i
- d_i The demand of flow i
- r_{iv} Monitoring of flow i by an IDS device of class v
- u_{jl} The capacity of link (j, l)
- x_{ij}^{A}, x_{ji}^{E} Access and egress flow connectivity
- h_{ik} The membership of i and k to the same resilient group

Variables

- o_{iv} Exclusion of flow i from monitoring by device of class v
- q^i_{jv} The monitoring of flow i in RN j by an IDS of class v
- s_{ji} The routing of flow i by RN j
- t^i_{jl} The routing of flow i on link (j, l)
- w^A_{ij}, w^E_{ji} The selection of access/egress RN j
- z_j The selection of RN j

Optimization Problem

- Minimize

$$F^* = \min \sum_{j,l \in J, i \in I} c^L_{jl} d_i t^i_{jl} + \sum_{i \in I, j \in J, v \in V} c^V_{jv} d_i q^i_{jv} + \sum_{i \in I, v \in V} c^P_i o_{iv},$$

- Multi-commodity flow conservation constraints

$$\sum_{j \in J} w^A_{ij} = 1, \sum_{j \in J} w^E_{ji} = 1, \quad \forall i \in I$$

$$w^A_{ij} \leq x^A_{ij} z_j, w^E_{ji} \leq x^E_{ji} z_j, \quad \forall i \in I, j \in J$$

$$w^A_{ij} - w^E_{ji} - \sum_{l \in J} \left(t^i_{jl} - t^i_{lj} \right) = 0, \quad \forall j \in J, i \in$$

- Bandwidth capacity constraints

$$\sum_{i \in I} d_i t^i_{jl} \leq u_{jl} z_j, \sum_{i \in I} d_i t^i_{jl} \leq u_{jl} z_l \quad \forall j, l \in J$$

Optimization Problem

- Selection of routing node and detection device on RN

$$\alpha s_{ji} \geq w_{ij}^{A} + w_{ji}^{E} + \sum_{l \in J} t_{jl}^{i}, \quad \forall i \in I, j \in J$$

$$s_{ji} \leq w_{ij}^{A} + w_{ji}^{E} + \sum_{l \in J} t_{jl}^{i}, \quad \forall i \in I, j \in J$$

$$q_{jv}^{i} \leq r_{iv} s_{ji}, \quad \forall j \in J, v \in V, i \in I$$

- Selection of distinct routing nodes for flows in the same resilient group

$$\sum_{k \in I, l \in J} (t_{jl}^{k} + t_{lj}^{k}) h_{ik} \leq 1, \qquad \forall i \in I, j \in J$$

Heuristic to Solve the IDS Network Design Problems

- Adopts the column-generation model
- Generate the set of the selected paths (P) between all flow's possible access and egress end-points
- Solve the minimal cost optimization subproblem
- Solve the IDS distribution optimization sub-problem

Theorem

The optimal cost of INDP is equal to the optimal cost of H-INDP, provided that the set of paths P_F resulting from the optimal selection of links (t^i_{jl}) in INDP is a subset of P and the cost of detection devices is independent from the location of RNs.

Path generation

Algorithm 1 Path Generator Algorithm

$P = \emptyset$;
for each $i \in I, j, j' \in J$ **do**
  if $x_{ij}^{A} <> 0$ and $x_{j'i}^{E} <> 0$ **then**
    $P' =$ @GeneratePaths(j, j');
    $\gamma_i^p = 1, \forall p \in P'$;
    $c_{ip}^{F} = c_{ij}^{A} + c_{j'i}^{E}, \forall p \in P'$;
    $P = P \cup P'$;
  end if
end for
$\delta_{jl}^{p} = 1$ if $(j, l) \vdash p = \texttt{True}, \forall j, l \in J, p \in P$.

Minimal Cost Optimization Subproblem

$$H_P^* = \min \sum_{j,l \in J, i \in I, p \in P} c_{jl}^L d_i \delta_{jl}^p y_i^p + \sum_{i \in I, p \in P} c_{ip}^F d_i \gamma_i^p y_i^p,$$

$$\sum_{i \in I, p \in P} d_i \delta_{jl}^p y_i^p \leq u_{jl}, \quad \forall j, l \in J$$

$$\sum_{p \in P} y_i^p = 1, \forall i \in I, \; y_i^p \leq \gamma_i^p, \forall i \in I, p \in P$$

$$\sum_{k \in I, l \in J, p \in P} (\delta_{jl}^p + \delta_{lj}^p) h_{ik} y_k^p \leq 1, \qquad \forall i \in I, j \in J$$

with the binary variable y_i^p with value 1 if flow i is routed on path $p \in P$, the binary parameter δ_{jl}^p with value 1 if path p contains the link (j, l), the binary parameter γ_i^p with value 1 if flow i can be routed on path p

IDS Device Distribution Optimization Subproblem

In previous step the optimal assignment of flows to paths $\hat{y}_i^p$, the sub-set of selected CRNs $J^S \subset J$, the sub-set of selected paths $P^S \subset P$, and the optimal selection of CRNs ζ_j^p was determined.

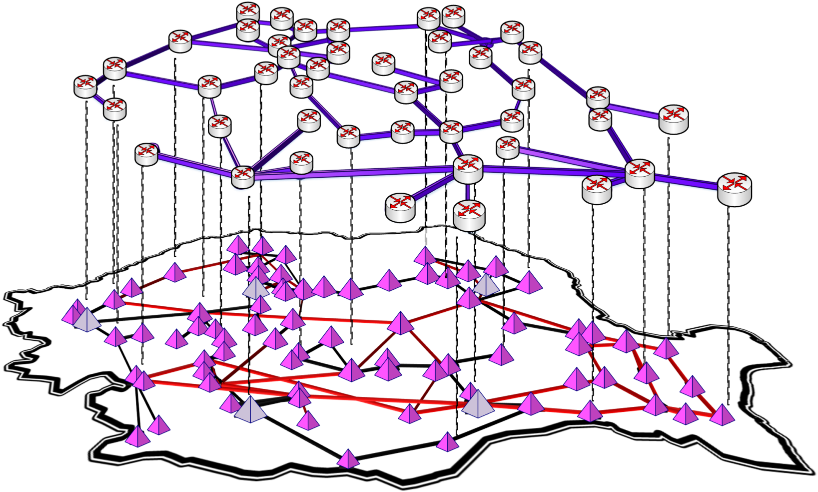
$$H_D^* = \min \sum_{i \in I, j \in J^S, v \in V} c_{jv}^V d_i q_{jv}^i + \sum_{i \in I, v \in V} c_i^P o_{iv},$$

$$q_{jv}^i \leq r_{iv} \sum_{p \in P^S} \hat{y}_i^p \zeta_j^p, \quad \forall i \in I, j \in J^S, v \in V$$

$$r_{iv} (\hat{y}_i^p - \sum_{j \in J^S} \zeta_j^p q_{jv}^i) \leq \hat{y}_i^p o_{iv}, \forall p \in P^S, i \in I, v \in V$$

Experimental network with 134 node, 176 flow

Secondary communication network (Romanian Educational Network - RoEduNet)



Physical infrastructure (Romanian 400kV and 220kV transmission network) and primary communication network

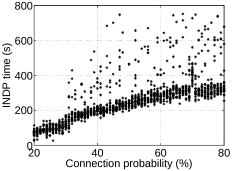
Table: INDP and H-INDP: Initial Setting

INDP		H-INDP^d				
Time [s]	Cost [MU]	Depth	$\|P\|$	Time [s]	Cost [MU]	Gap [%]
919.3	567595	6	1795	1.3	567595	0
		7	3972	2.7	567595	0
		8	8623	6.02	567595	0

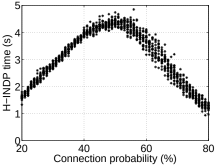
Table: INDP and H-INDP: Random Cost Distribution

INDP		H-INDP^d				
Time [s]	Cost [MU]	Depth	$\|P\|$	Time [s]	Cost [MU]	Gap [%]
337.9	3441304	6	1795	1.3	3923891	14.02
		7	3972	2.9	3895087	13.1
		8	8623	7.1	3890722	13.05

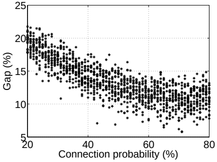
Computation time and gap with synthetic data



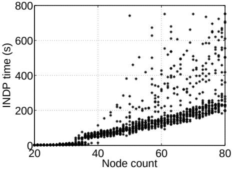
INDP.



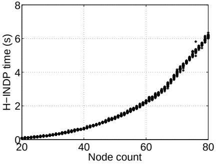
H-INDP^s.



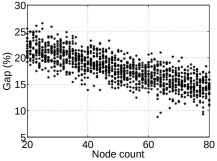
H-INDP^s.



INDP.



H-INDP^s.



H-INDP^s.

References

- B. Genge, P. Haller, I. Kiss: Big Data Processing to Detect Abnormal Behavior in Smart Grids, 1st EAI International Conference on Smart Grid Inspired Future Technologies, MAY 19-20, 2016, Liverpool, Great Britain.
- B. Genge, P. Haller, I. Kiss: *Cyber Security-Aware Network Design of Industrial Control Systems*, IEEE Systems Journal, IEEE Systems Council, 2015 (Accepted).
- P. Haller, B. Genge: Using Sensitivity Analysis and Cross-Association for the Design of Intrusion Detection Systems in Industrial Cyber-Physical Systems, IEEE Access, vol. 5, pp. 9336-9347, 2017.
- B. Genge, P. Haller, C.D. Dumitru, C. Enachescu: Designing Optimal and Resilient Intrusion Detection Architectures for Smart Grids, IEEE Transactions on Smart Grid, 2017.