Estimation of distribution for anomaly detection in Water Distribution Networks

Séminaire Doctorant ‘Détection d’attaques’ - 21/06/2024

Côme Frappé - - Vialatoux

ICube, UMR 7357, Université de Strasbourg, CNRS, 67000 Strasbourg, France

come.frappe-vialatoux@etu.unistra.fr

Estimation of distribution for anomaly detection in Water Distribution Networks

Plan

- Water distribution network
- Anomaly detection
- Estimation of distribution Algorithms (EDA)

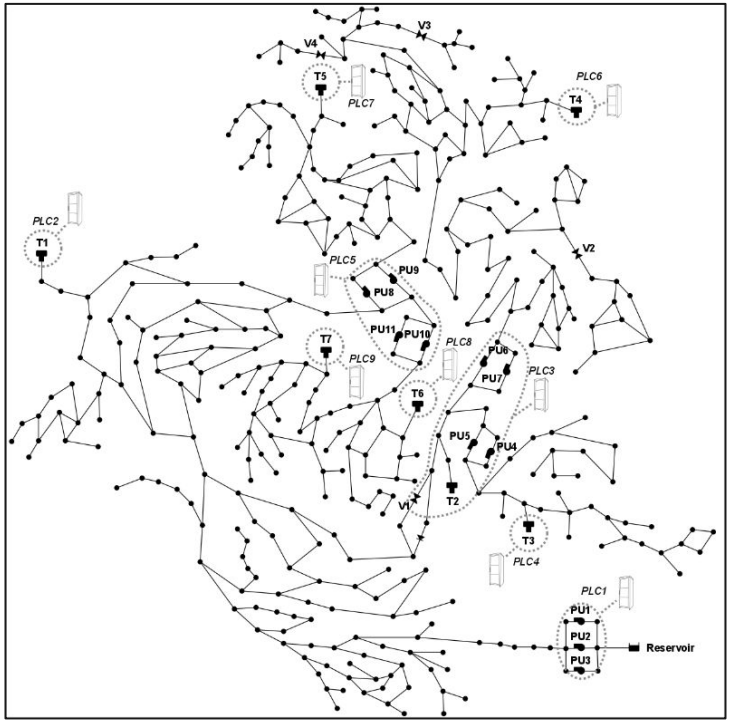
Plan

- Water distribution network
  - ➢ Context
- Anomaly detection
  - ➢ Problem
- Estimation of distribution Algorithms (EDA)
  - ➢ Solution

Water distribution networks

- Composed of Pipes, Tanks, Valves, Pumps, etc.
- Mesh an area
- provide water to endpoints

→ Essential to the society



Water distribution network from the BATADAL[1] dataset

Water distribution networks

Not operated manually !

Modernization → automation → connected devices

- Better monitoring
- Better adaptability
- Better control

Water distribution networks

Not operated manually !

Modernization → automation → connected devices

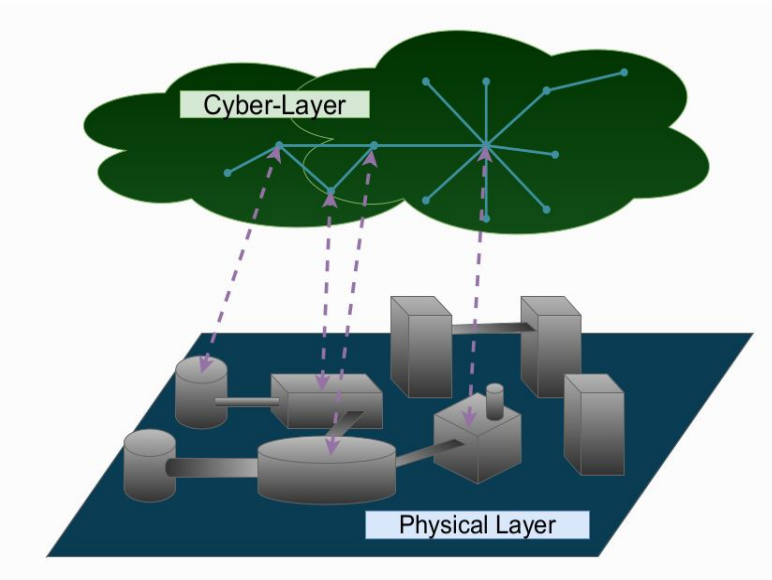
Increased Vulnerability to cyber-attack !

- Better monitoring
- Better adaptability
- Better control

Water distribution networks

Transition from Physical systems to **Cyber-physical systems**

Attack perimeter includes both layers



Anomaly Detection

- Detecting anomalies with network data
  - malicious
  - non-malicious (malfunctions)

2 main approaches

- Supervised
- Unsupervised

Anomaly Detection

- Detecting anomalies with network data
  - malicious
  - non-malicious (malfunctions)

2 main approaches

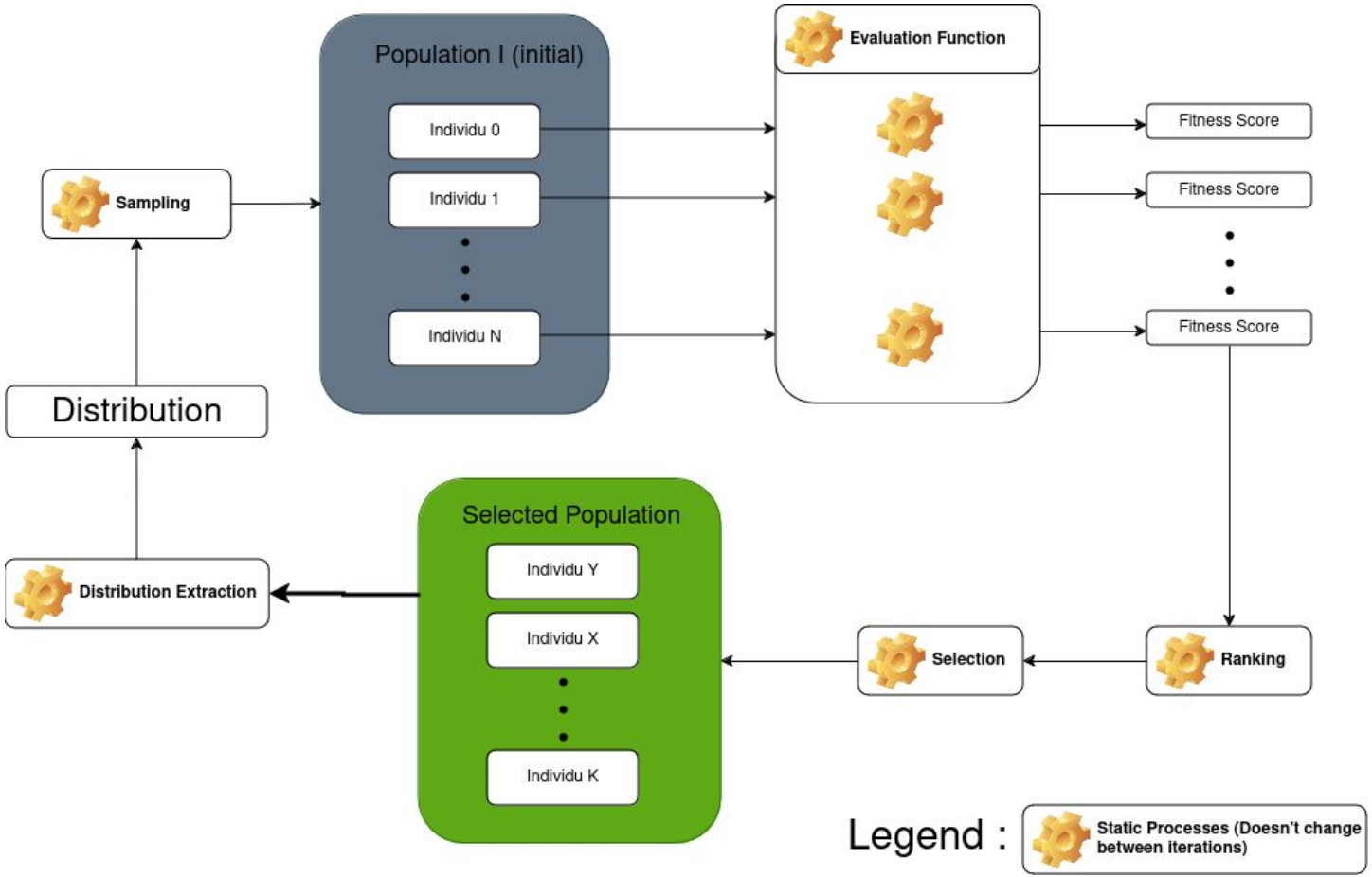
- Supervised ➢ Training data *with labeled attacks included*
- Unsupervised ➢ Training data *clean of attacks*

Estimation of Distribution algorithms

General description

- Family of evolutionary algorithms
- Population based
- **Extract and optimise a probability distribution**

EDA :


Population I (initial)
Individu 0
Individu 1
Individu N
Evaluation Function
Fitness Score
Fitness Score
Fitness Score
Sampling
Distribution
Distribution Extraction
Selected Population
Individu Y
Individu X
Individu K
Selection
Ranking
Legend :
Static Processes (Doesn't change between iterations)

Estimation of Distribution algorithms

Process for anomaly detection: Unsupervised

1. Extract the distribution of legitimate data → Reference
2. Extract the current data distribution
3. Compare the two distributions
    a. Close match = Normal
    b. Mismatch = Anomaly !

Estimation of Distribution algorithms

For anomaly Detection: Supervised

1. Extract the distribution of each label (including normal)
2. Extract the current data distribution
3. Find the closest distribution → Current state

Estimation of Distribution algorithms

Why it is a novel approach:

- EDA never used on network data
  - Mainly used on binary vectors
  - Only Theoretical work on generalisation to r-valued vectors [2]
  - Non-numerical data needs encoding

Estimation of Distribution algorithms

Why it is a novel approach:

- EDA never used for detection → Challenges
  - Express detection as an optimisation process
    - Score function representative of attack ‘level’
  - multiple proximity measures possible
    - divergence
    - distance

Estimation of Distribution algorithms

Expected advantages

- State of the system = Distribution
  - single object representation
- EDA is robust to noise
- EDA could identify informative features
- Potentially agnostic of data type
  - possible extension with physical data

Conclusion:

- Prospective work
  - Many open research questions
  - New approach for detection
  - Theoretical work + technical work (no library of EDA for detection !)

References

1.Taormina, R.; Galelli, S.; Tippenhauer, N.O.; Salomons, E.; Ostfeld, A.; Eliades, D.G.; Aghashahi, M.; Sundararajan, R.; Pourahmadi, M.; Banks, M.K.; et al. Battle of the Attack Detection Algorithms: Disclosing Cyber Attacks on Water Distribution Networks. *Journal of Water Resources Planning and Management* **2018**, *144*, 04018048, doi:10.1061/(ASCE)WR.1943-5452.0000969.

2.Ben Jedidia, F.; Doerr, B.; Krejca, M.S. Estimation-of-Distribution Algorithms for Multi-Valued Decision Variables. In Proceedings of the Proceedings of the Genetic and Evolutionary Computation Conference; Association for Computing Machinery: New York, NY, USA, July 12 2023; pp. 230–238.

Bonus: Pelikan, M.; Hauschild, M.W.; Lobo, F.G. Estimation of Distribution Algorithms. In *Springer Handbook of Computational Intelligence*; Kacprzyk, J., Pedrycz, W., Eds.; Springer: Berlin, Heidelberg, 2015; pp. 899–928 ISBN 978-3-662-43505-2.