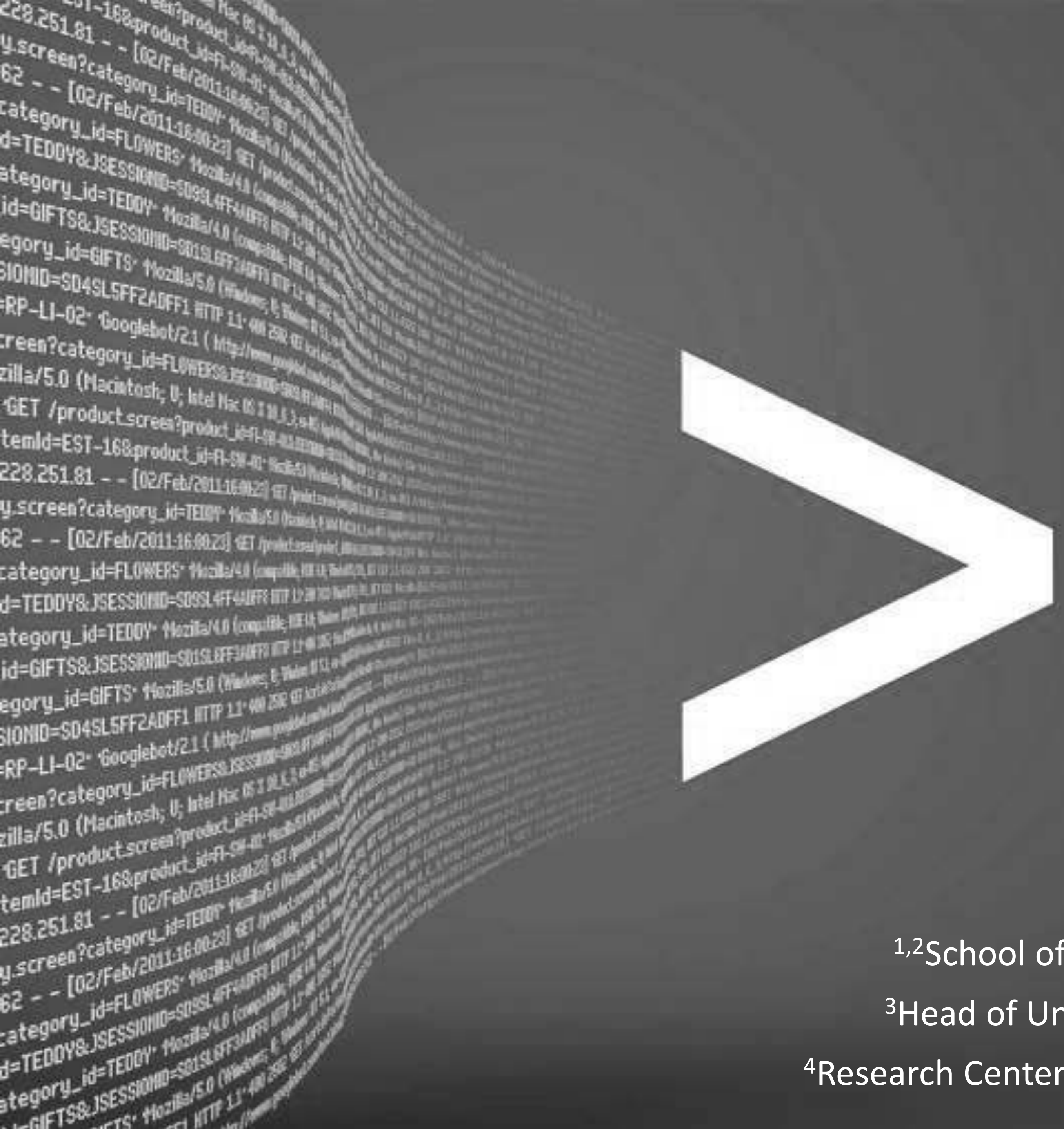


Cyber-Typhon: An Online Multi-Task Anomaly Detection Framework

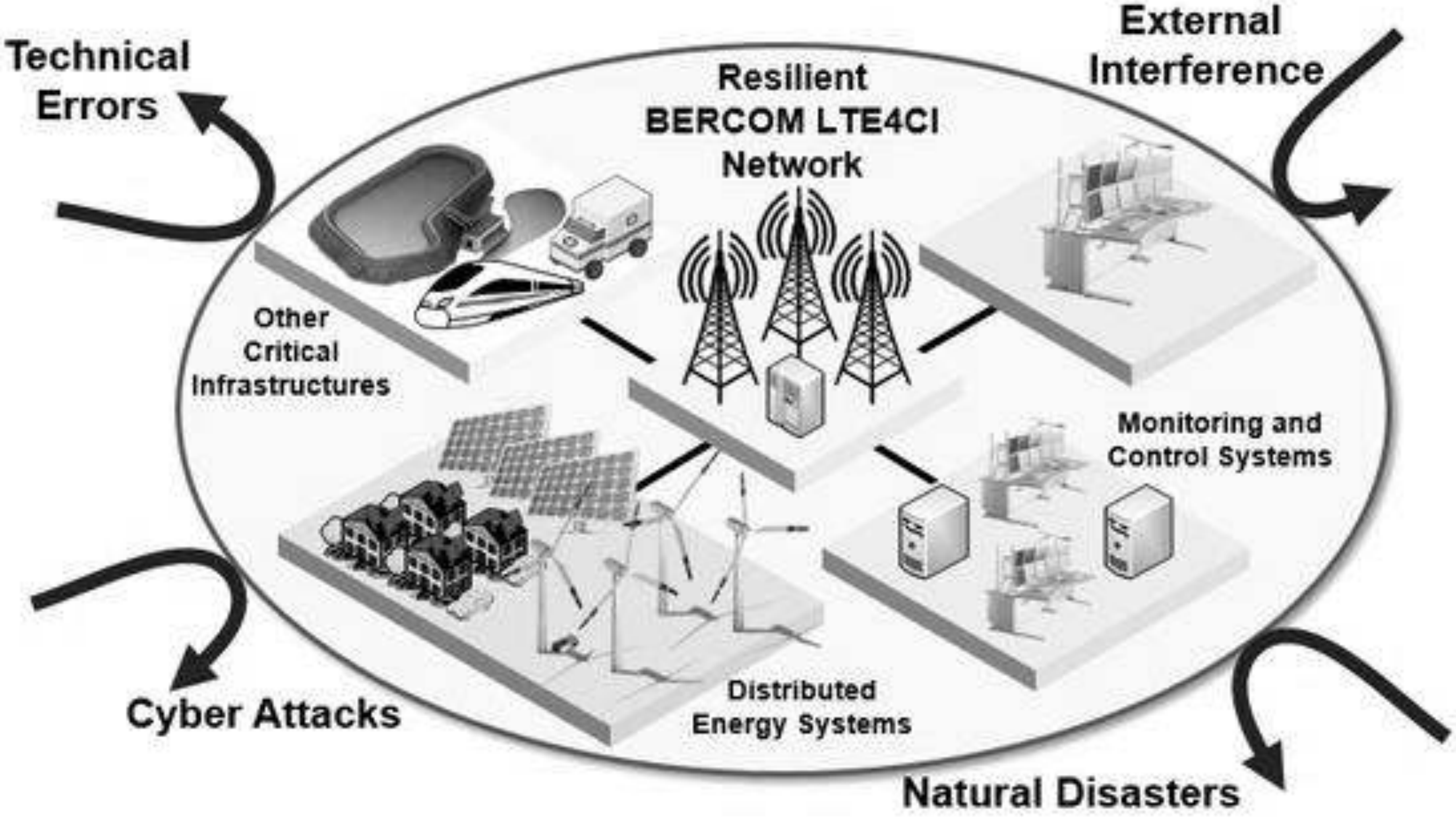
K. Demertzis[1], L. Iliadis[2], P. Kikiras[3], N. Tziritas[4]

[1,2]School of Civil Engineering, Democritus University of Thrace, Greece

[3]Head of Unit Innovative Research, European Defense Agency, Belgium

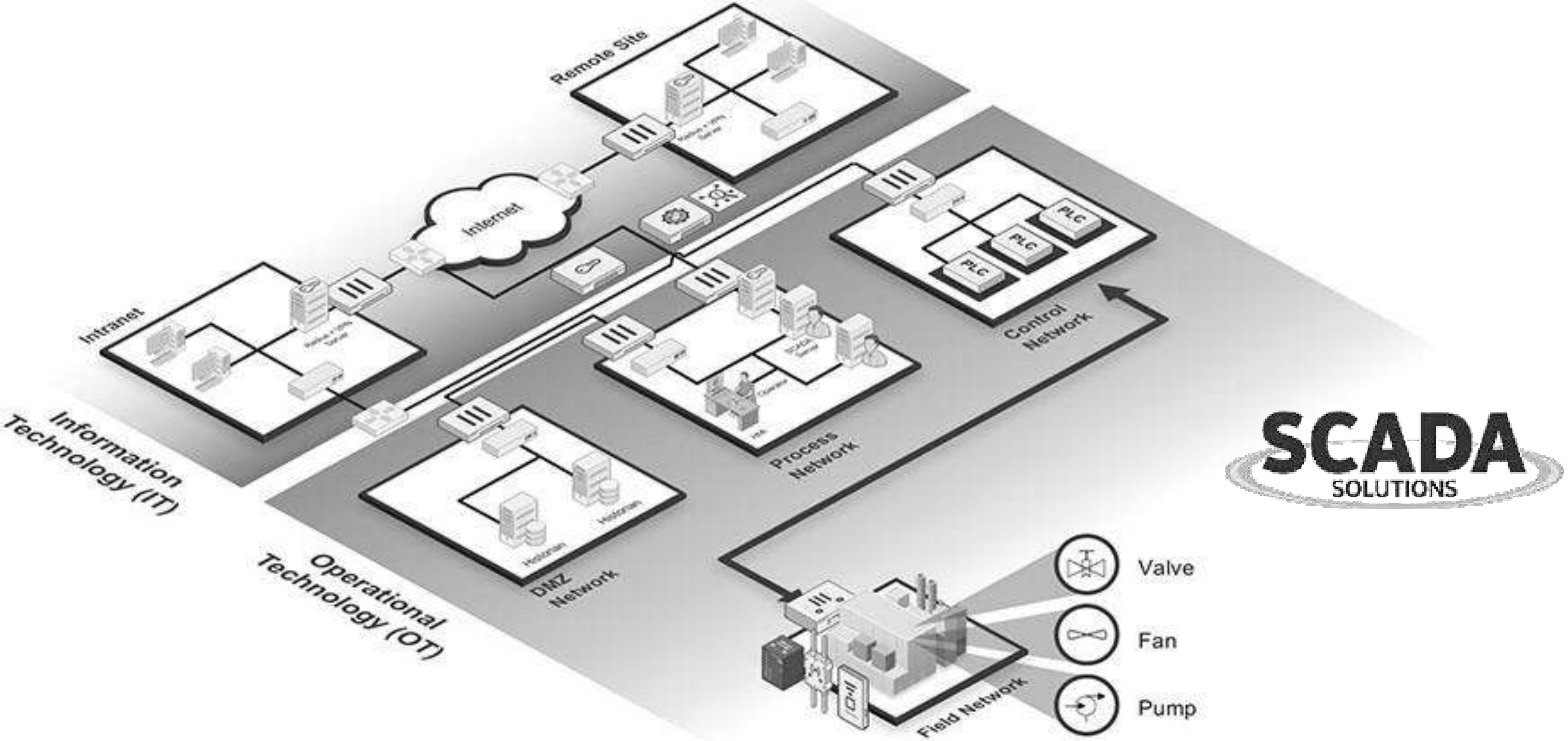
[4]Research Center for Cloud Computing, Chinese Academy of Sciences, China

Cybersecurity Protection of Critical Infrastructures



Technical Errors
Resilient BERCOM LTE4CI Network
External Interference
Other Critical Infrastructures
Monitoring and Control Systems
Distributed Energy Systems
Cyber Attacks
Natural Disasters

Cybersecurity Protection of Critical Infrastructures

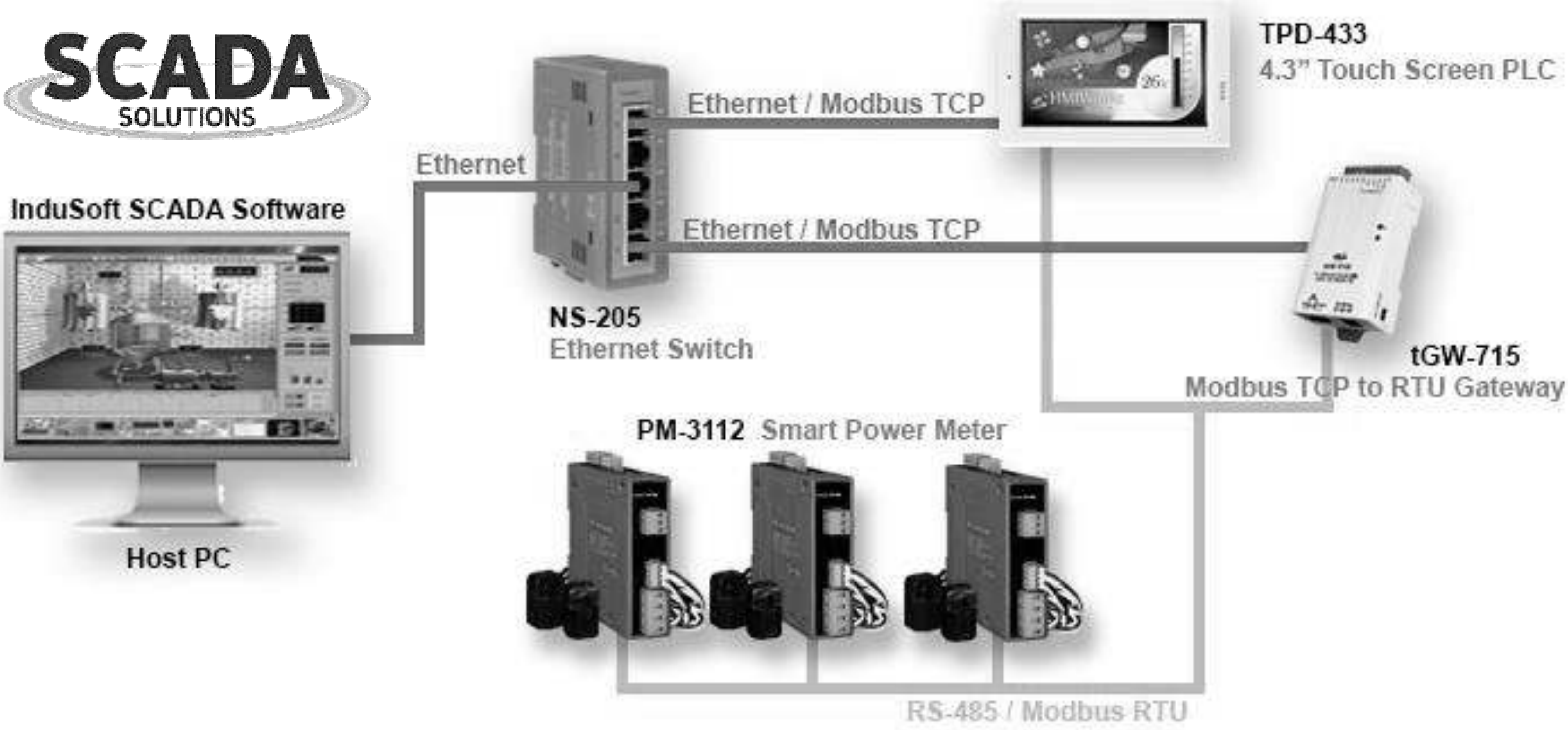


Cybersecurity Protection of Critical Infrastructures

- SCADA Systems and Distribution Control Systems:
  - ancillary systems that are the basis of most integrated ICS architectures,
  - programmable logic controllers (PLC),
  - remote terminal units (RTU),
  - intelligent electrical device (IED),
  - basic process controllers (BPCS),
  - safety instrumented systems (SIS) and
  - operator panels.

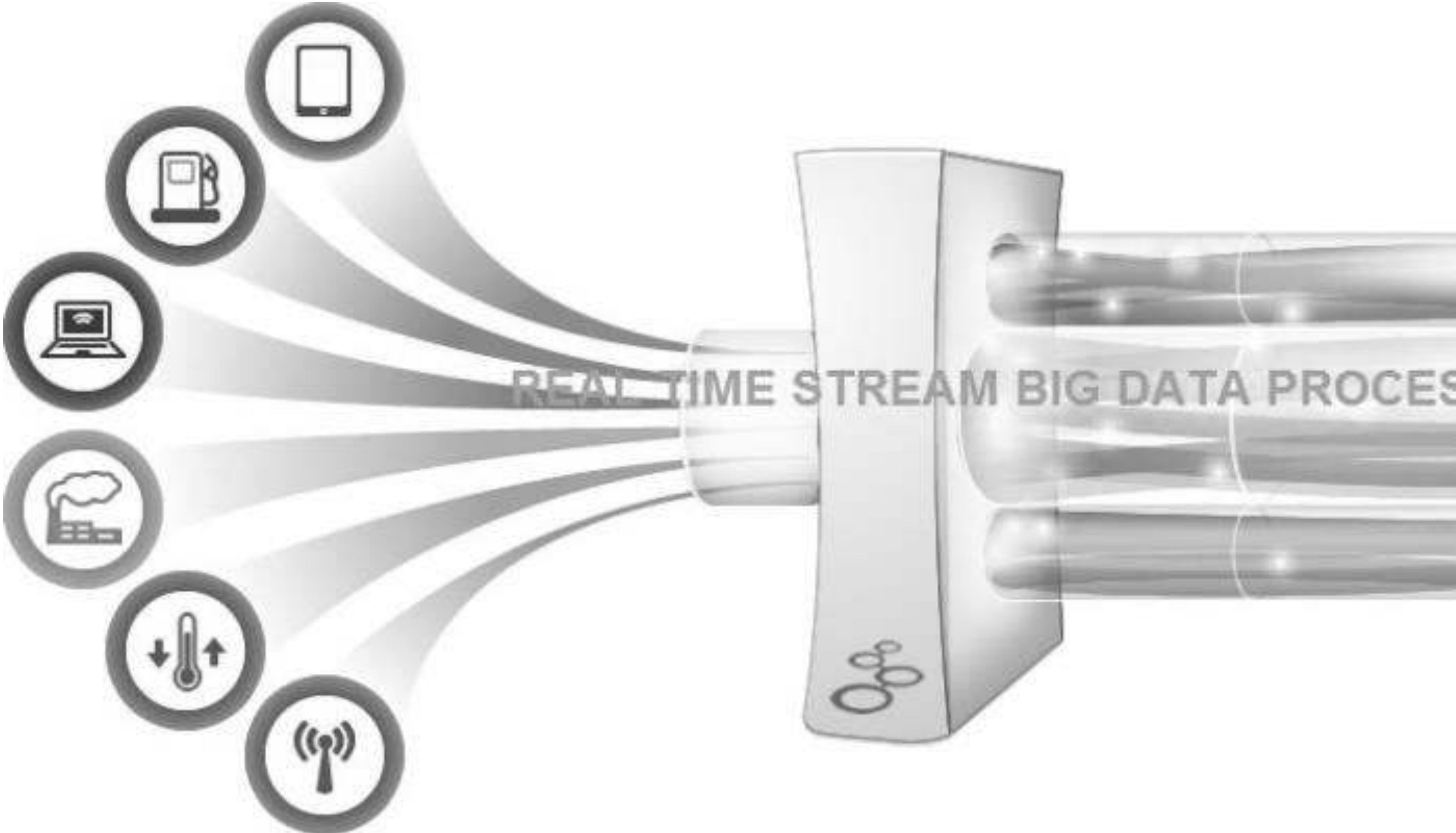
SCADA
SOLUTIONS

Cybersecurity Protection of Critical Infrastructures

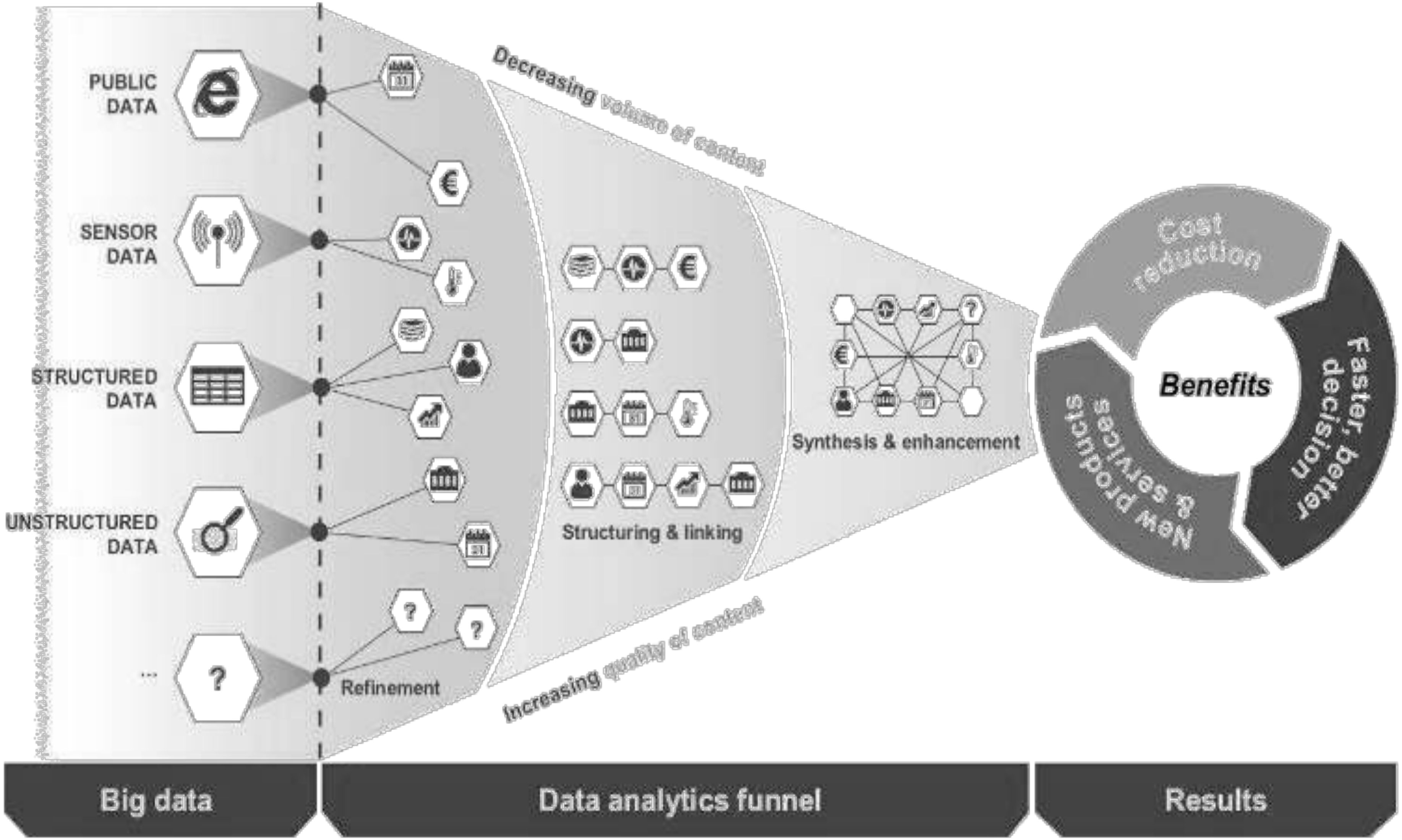


All data originates in real-time!

Real Time Big Data Stream Processing



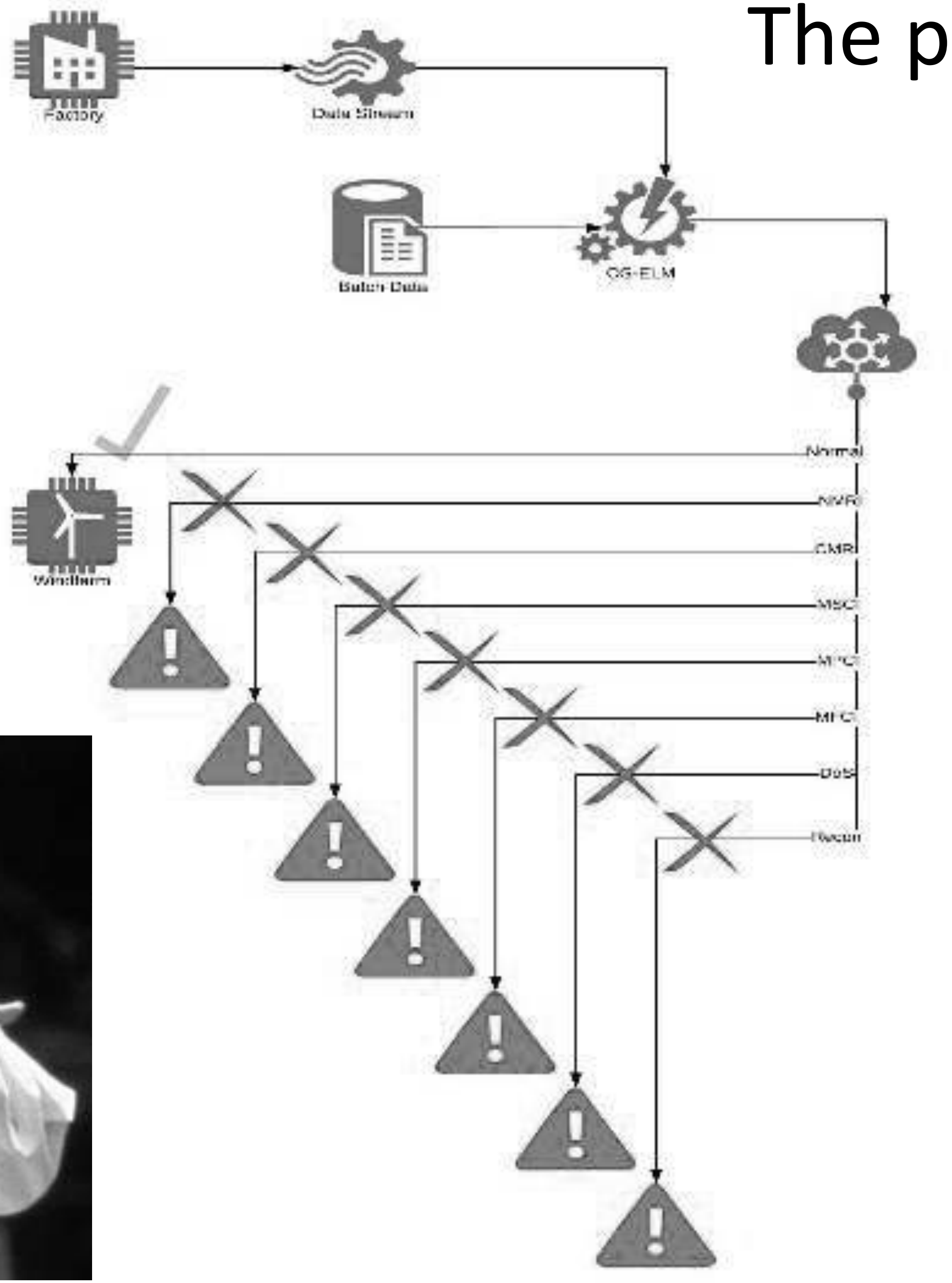
Large-Scale Data Analytics

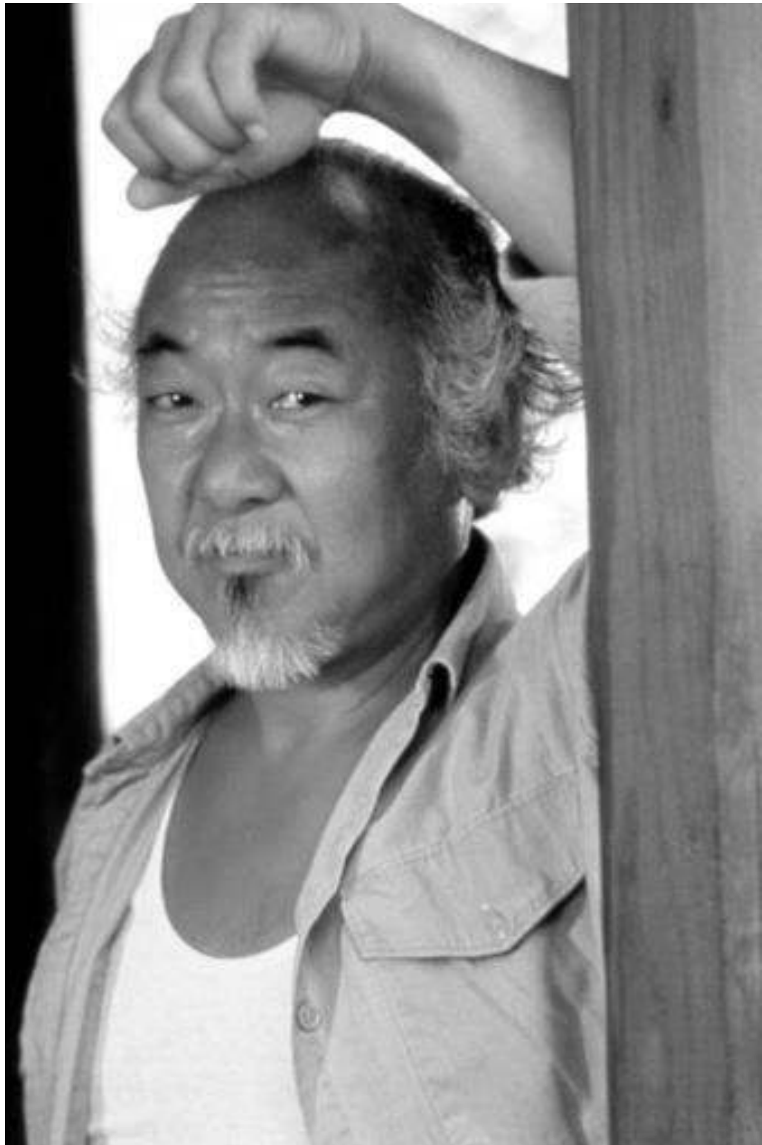


Multi-Task Learning

- The following approaches are characteristic cases of MTL:
  - **Task grouping and overlapping**
  - **Exploiting unrelated tasks**
  - **Transfer of knowledge**
  - **Group online adaptive learning**

The proposed Cyber-Typhon Framework

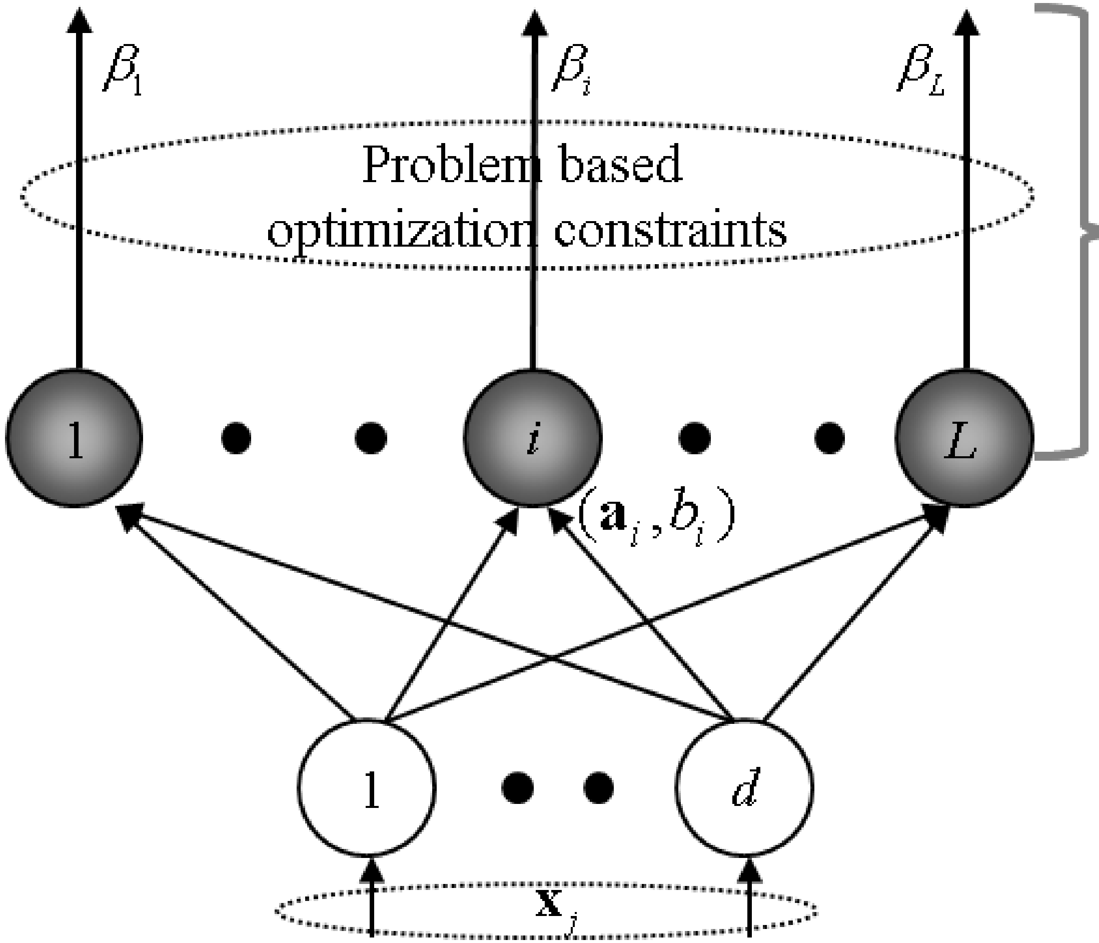




The proposed Cyber-Typhon Framework

- The Cyber-Typhon initially exports features related to network traffic, which are used as input to an OS-ELM neural network.
- The OS-ELM has been trained with proper data, in order to be able either to classify traffic as normal or (in the opposite case) to identify the threat or the attack type.
- If the network traffic is normal further communication is allowed.
- In the opposite case, the type of anomaly is determined and the data flow is redirected to a proper absolutely specialized and dedicated RBM.
- If the first RBM does not recognize the specific anomaly for which it is specialized, the data is redirected to the next RBM responsible for the detection of another anomaly and so on till the successful identification is achieved.
- If detection cannot be done by any of the trained RBM (which are as many as the types of the known anomalies) the network flow data return to the initial OS-ELM, which can perform online sequential learning (thus, the classification effort can be re-adjusted).

OS-ELM



Feature learning
Clustering
Regression
Classification

L Random Hidden Neurons (which need not be algebraic sum based) or other ELM feature mappings. Different type of output functions could be used in different neurons:

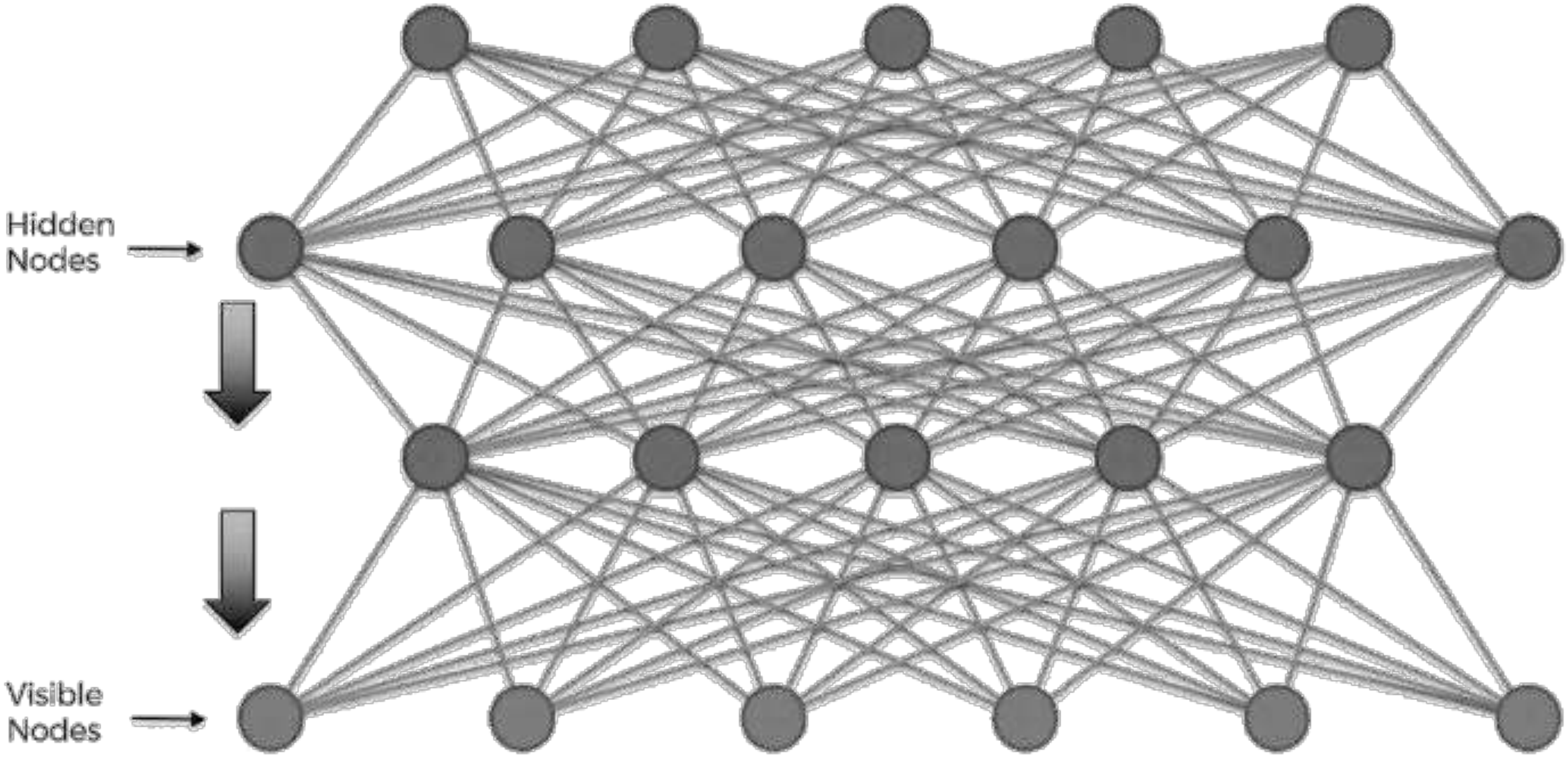
$$h_i(\mathbf{x}) = G_i(\mathbf{a}_i, b_i, \mathbf{x})$$

d Input Nodes

Online Sequential ELM

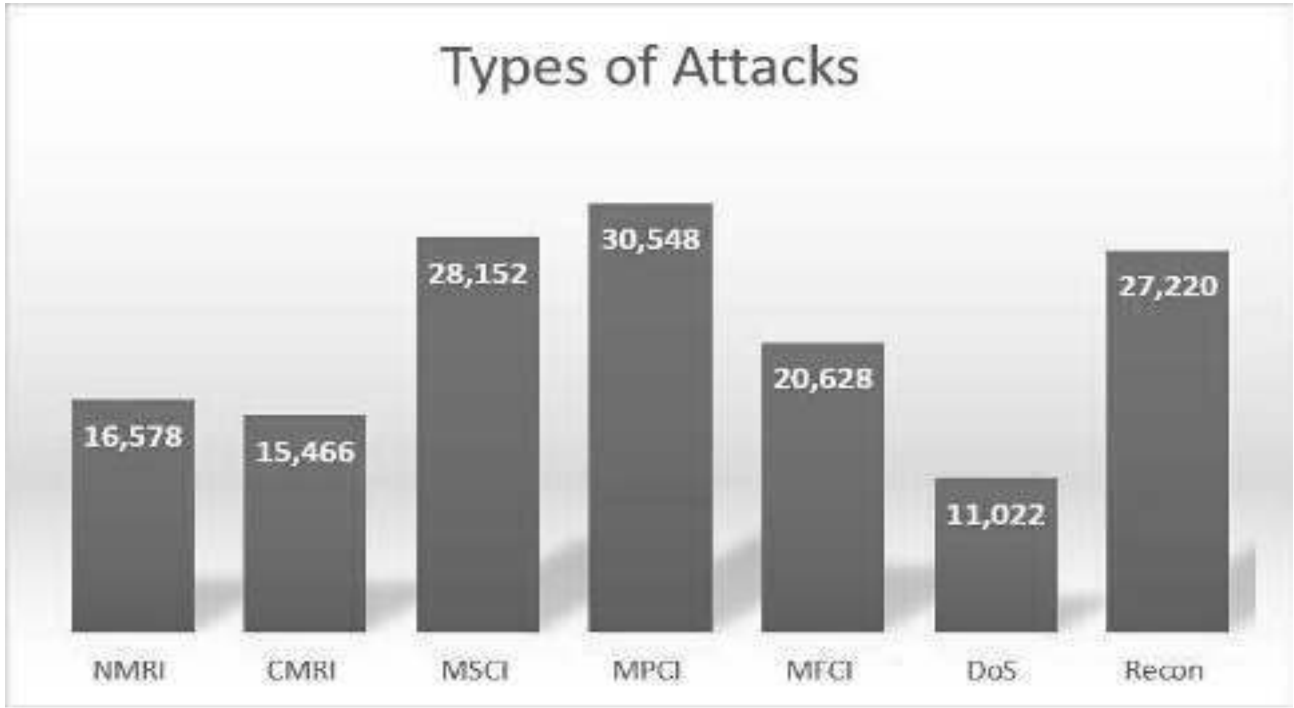
- used over a sliding data window,
- can learn the sequential training observations online at arbitrary length (one by one or chunk by chunk) with fixed or varying length and discard the data for which the training has already been done,
- it has no prior knowledge about the amount of the observations which will be presented,
- do not require retraining whenever a new data is received,
- as soon as the learning procedure for the arrived observations is completed, the data is discarded.

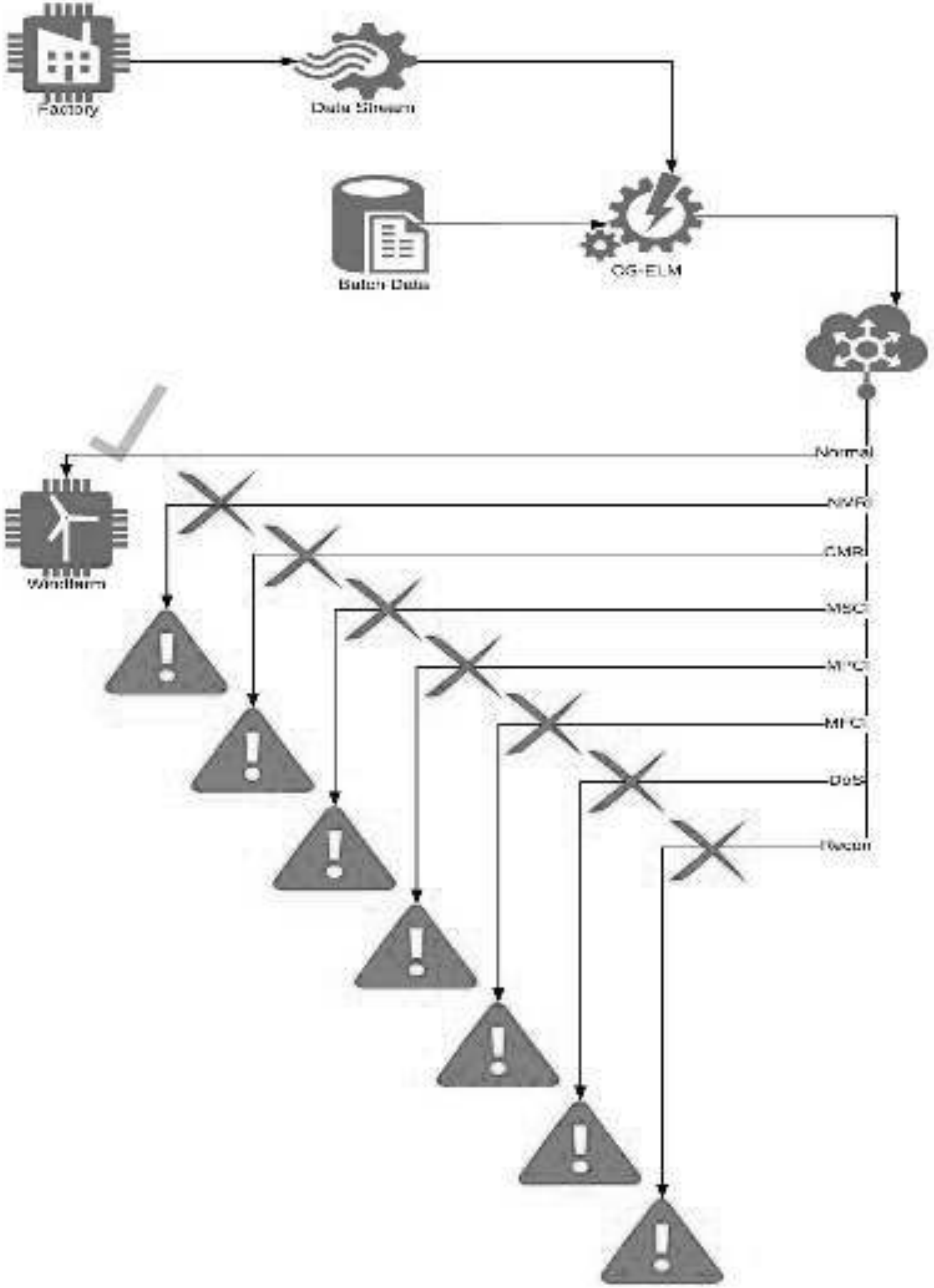
RBM



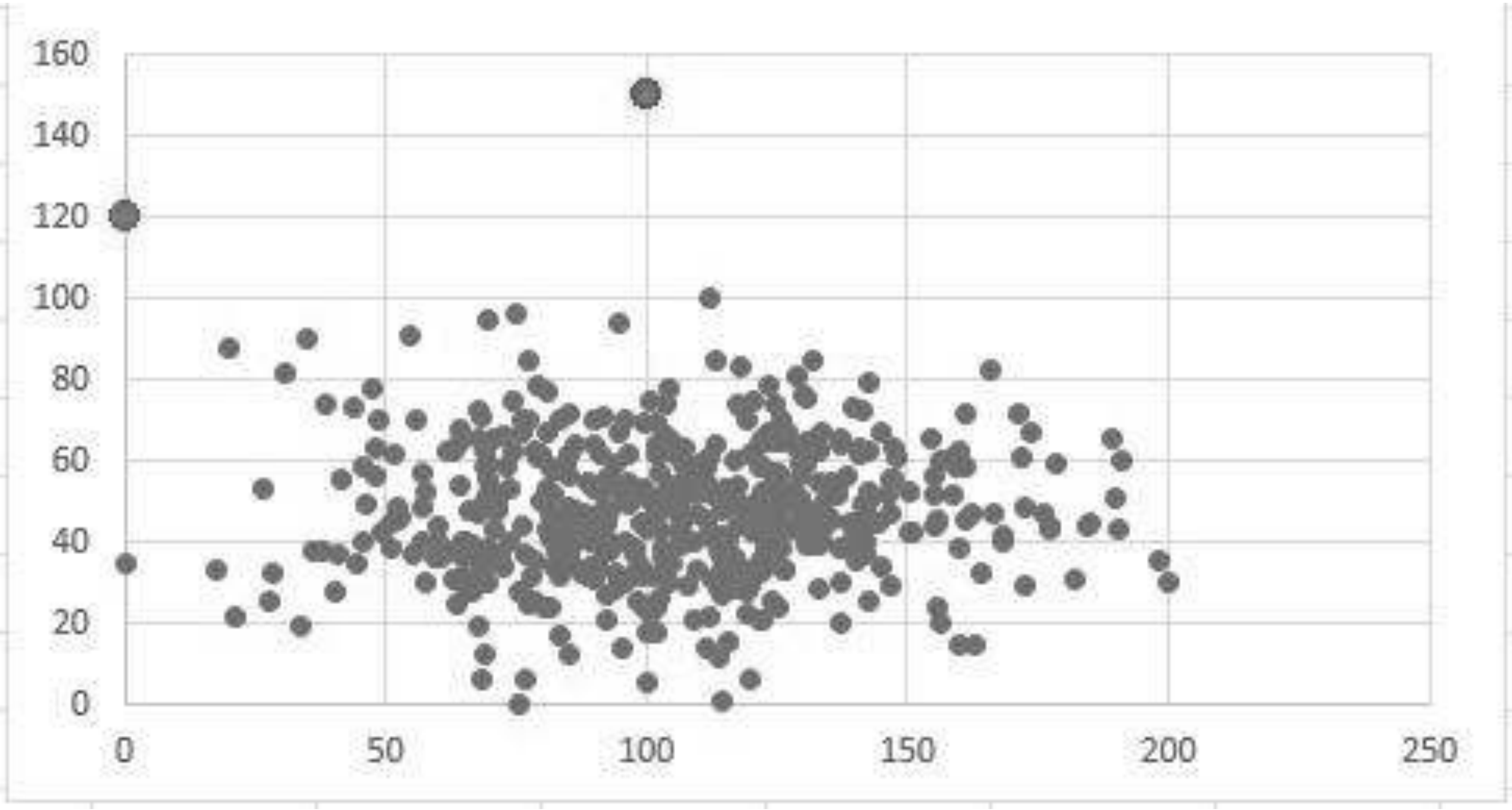
The proposed Cyber-Typhon Framework

- The Cyber-Typhon there are 7 RBMs, as many as the types of attacks, where each one of them has been trained to perform One-Class Classification in order to exclusively recognize one specific network attack.

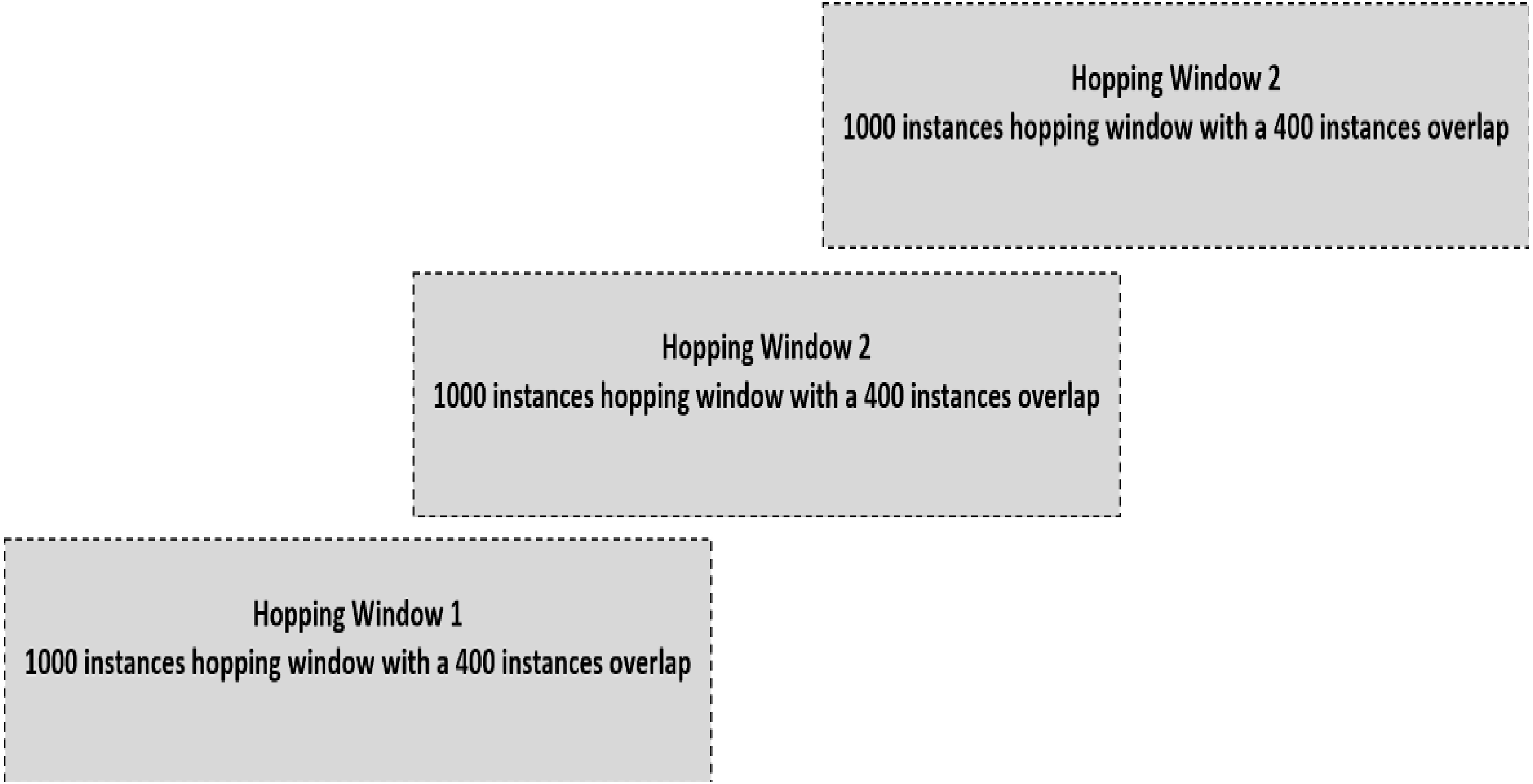




OCC



MTL



Hopping Window 2

1000 instances hopping window with a 400 instances overlap

Hopping Window 2

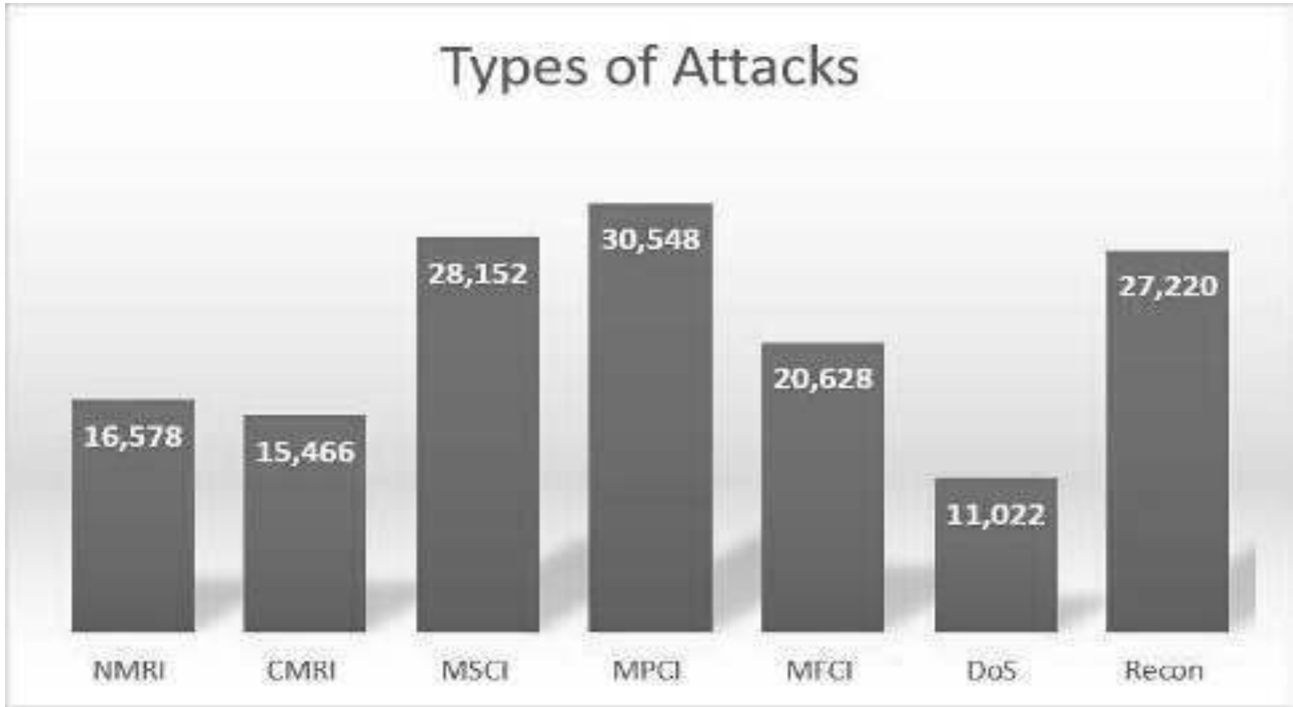
1000 instances hopping window with a 400 instances overlap

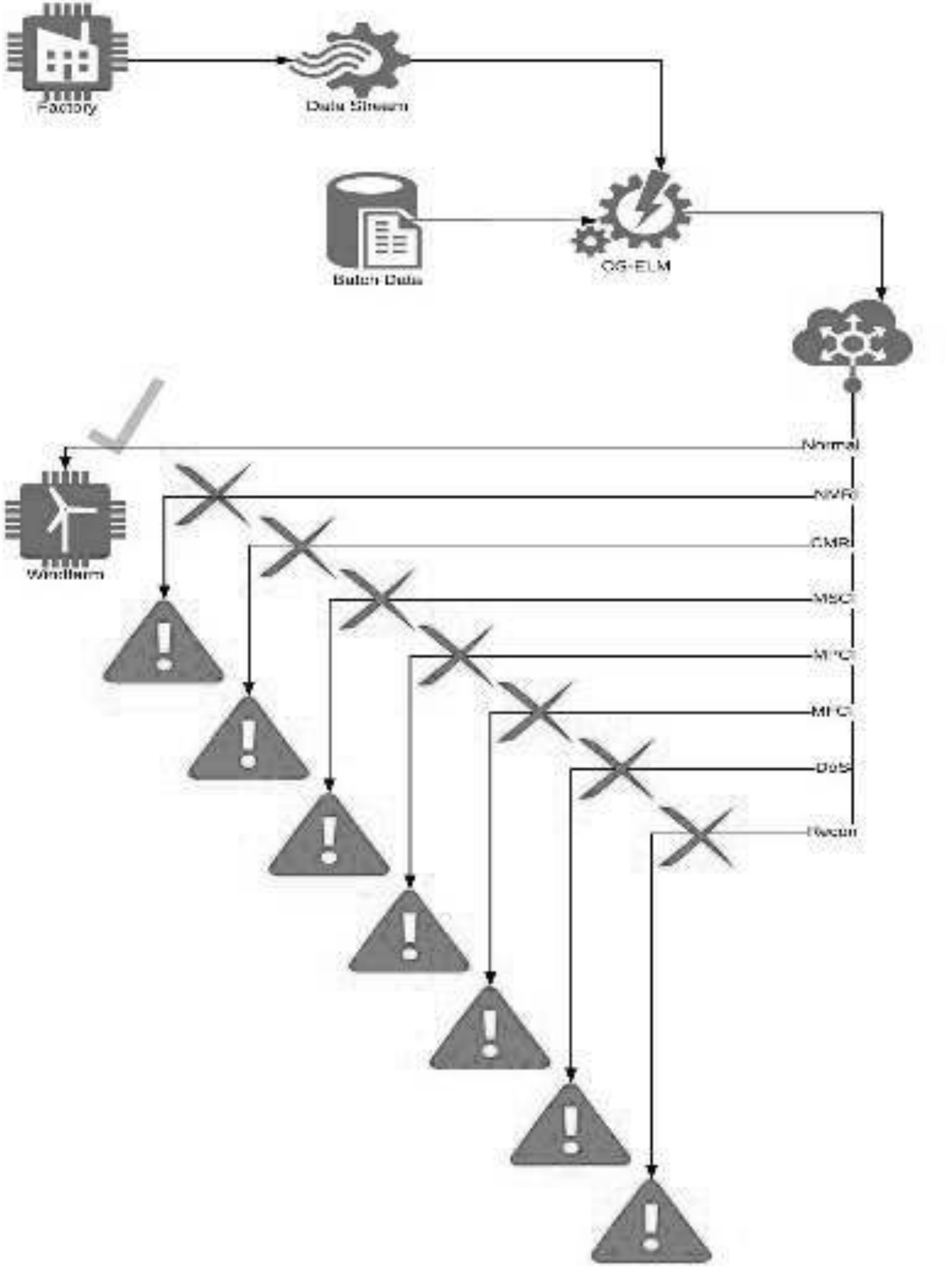
Hopping Window 1

1000 instances hopping window with a 400 instances overlap

The proposed Cyber-Typhon Framework

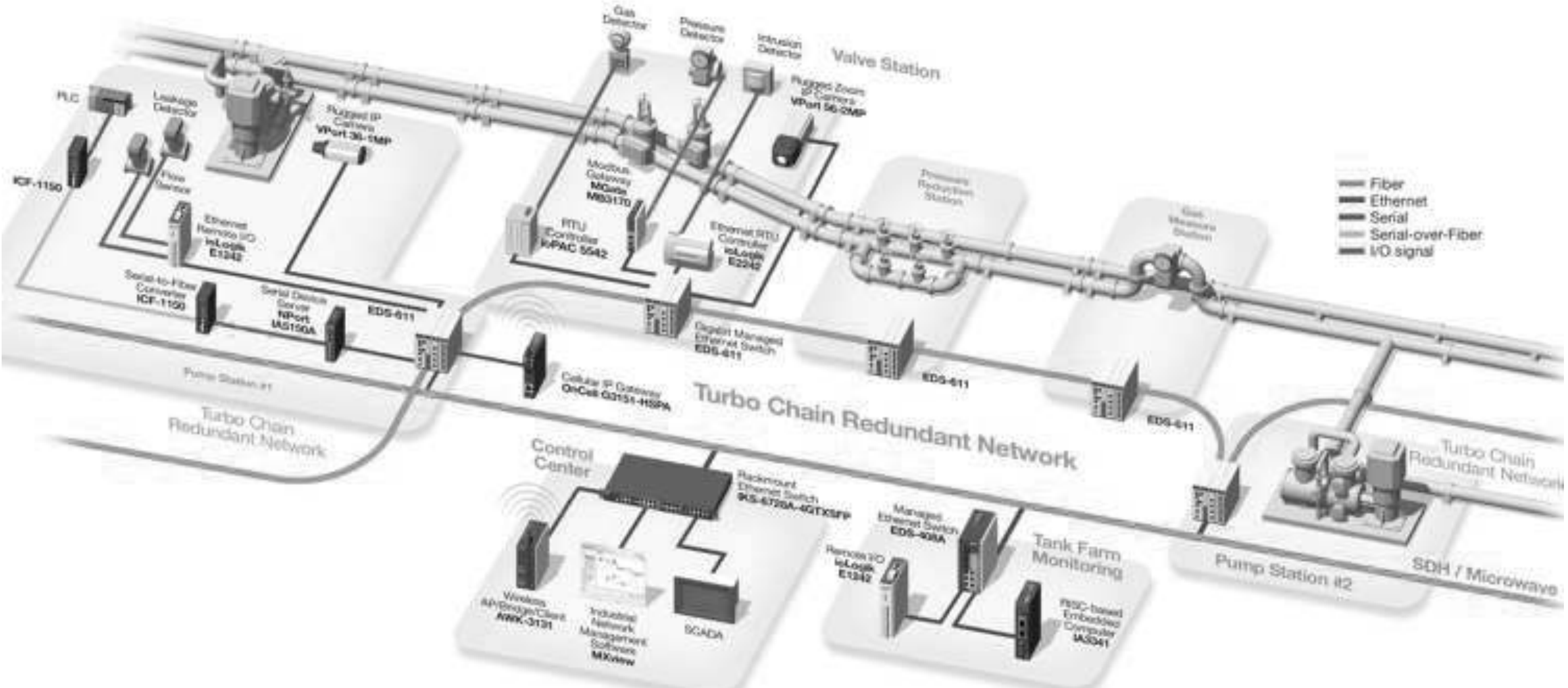
- The Cyber-Typhon there are 7 RBMs, as many as the types of attacks, where each one of them has been trained to perform One-Class Classification in order to exclusively recognize one specific network attack.





DATASET

- The **gas_dataset** includes 26 independent features and **97,019** instances, from which ***61,156 normal*** and ***35,863 outliers***. The training of the algorithm was done with the gas_train_dataset that contains ***30,499 normal instances***, whereas the rest **30,657 normal instances and 35,863 outliers**, belong to the gas_test_dataset.

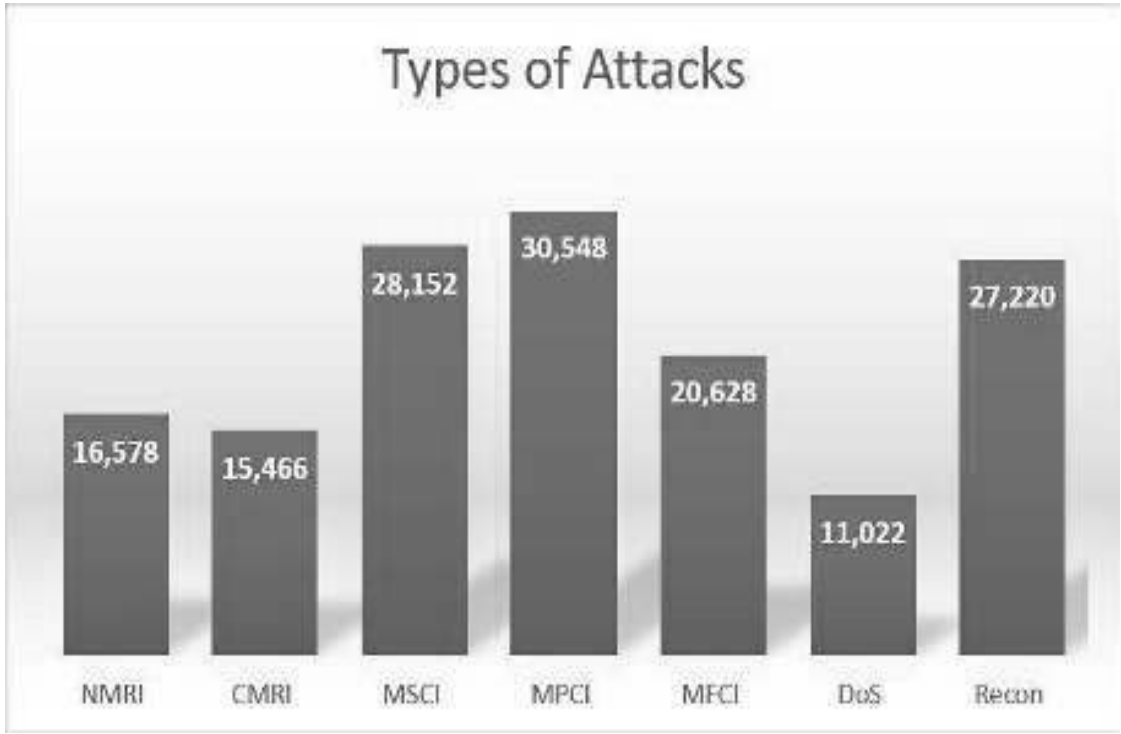


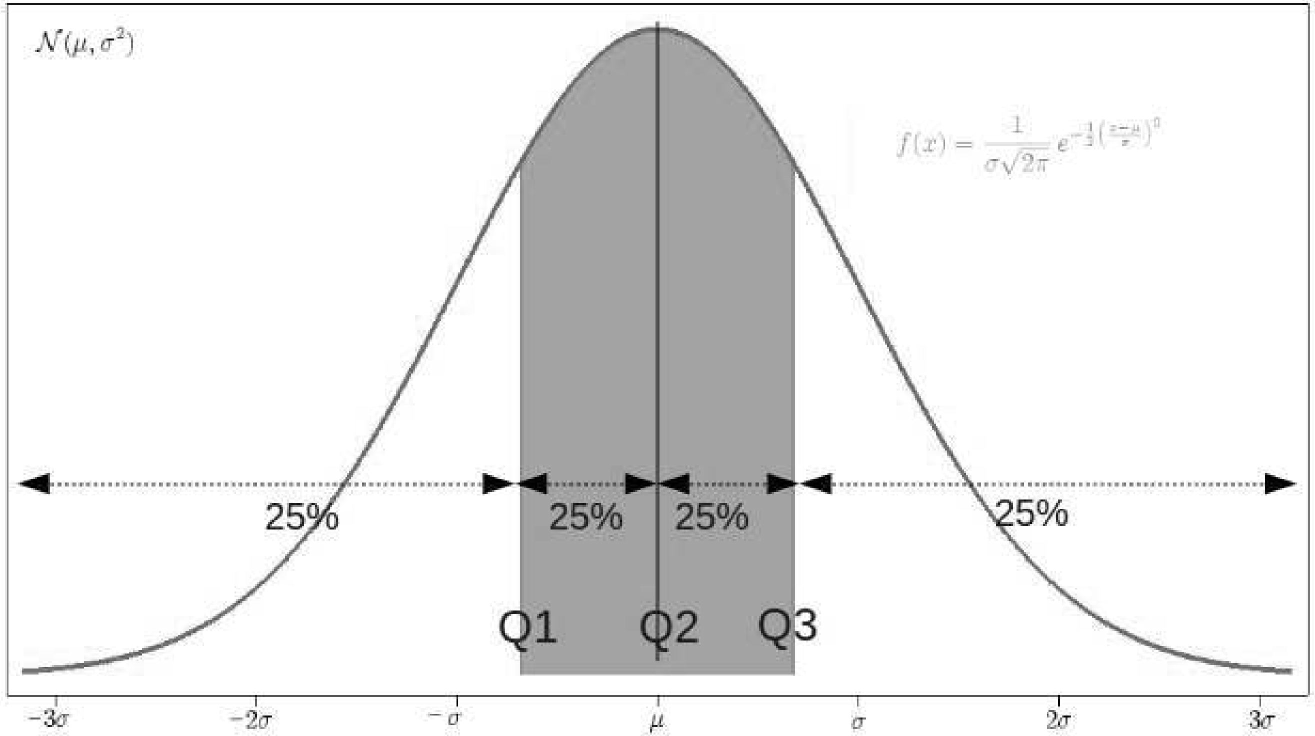
Real world data
iris & mtcars

IOT
Direction of traffic

DATASET

- The dataset is determined and normalized in the interval [-1,1] in order to phase the problem of prevalence of features with wider range over the ones with a narrower range, without being more important.
- Also, the outliers and the extreme values spotted were removed based on the Inter Quartile Range technique.





RESULTS

Table 1. Confusion Matrix of the OS-ELM

Normal	**NMRI**	**CMRI**	**MSCI**	**MPCI**	**MFCI**	**DoS**	**Recon**
59,826	428	93	289	453	2	65	0
632	**15,944**	**0**	**2**	0	0	0	0
40	0	**15,426**	**0**	0	0	0	0
264	0	0	**27,888**	0	0	0	0
503	0	0	0	**29,900**	125	20	0
2	0	0	0	157	**20,469**	0	0
139	0	0	1	24	0	**10,858**	0
0	0	0	0	0	0	0	**2,220**

RESULTS

Table 2. Classification Accuracy and Performance Metrics

Classifier	Fold	TA	RMSE	Precision	Recall	F-Score	AUC
OS-ELM	1st	98.51%	0.0548	0.980	0.980	0.9800	0.998
	2nd	98.63%	0.0541	0.990	0.990	0.9900	0.999
	3rd	97.96%	0.0482	0.976	0.976	0.9760	0.989
	4th	98.63%	0.0543	0.990	0.990	0.9900	0.996
	5th	98.98%	0.0578	0.989	0.989	0.9890	0.997
	6th	98.00%	0.0490	0.981	0.981	0.9810	0.995
	7th	98.60%	0.0549	0.986	0.986	0.9860	0.999
	8th	98.75%	0.0560	0.987	0.987	0.9870	0.999
	9th	98.28%	0.0567	0.986	0.986	0.9860	0.999
	10th	98.30%	0.0536	0.985	0.985	0.9850	0.999
	Avg	**98.46%**	**0.0539**	**0.985**	**0.985**	**0.985**	**0.997**

Future Work

- Proposals for the development and future improvements of this system, should focus on further optimizing the parameters of the RBMs used in order to achieve an even more efficient, accurate and quicker classification, capable of dividing even more precisely the boundaries between the situations of systems.
- It would be important to study the equation-extension of the proposed algorithm with meta-learning methods. This could further improve the anomaly detection process.
- Finally, the introduced model can employ adaptive learning in order to gain self-improvement potentials. This would automate 100% the whole process.

Any questions?