THE UNIVERSITY of EDINBURGH
informatics

Adaptive Clustering-based Malicious Traffic Classification at the Network Edge

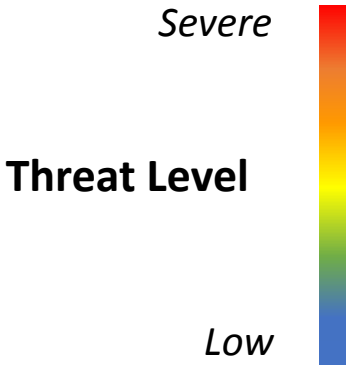
Alec F. Diallo, Dr. Paul Patras

Coseners - July, 2021

2 Network Intrusion Detection Systems

- Two main approaches: *Knowledge oriented* / *Data oriented*

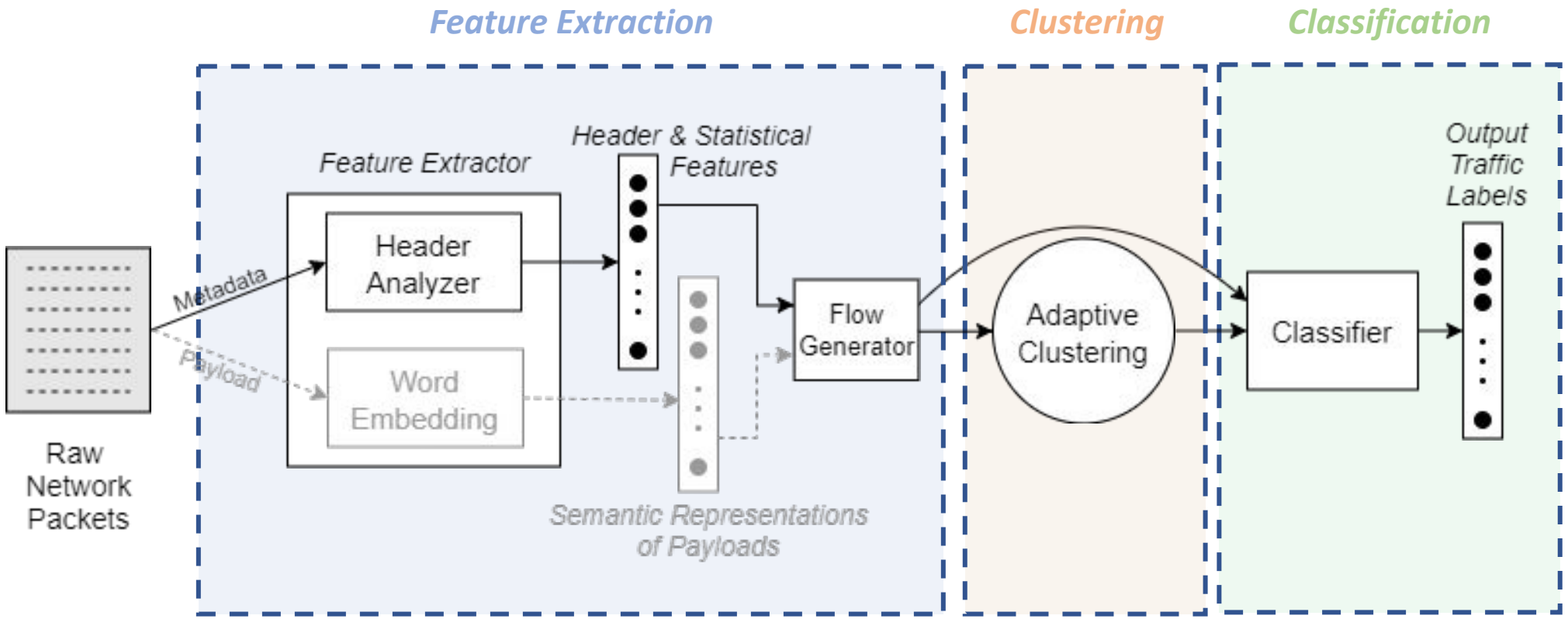
- Shortcomings of existing solutions:



- ★ Volume of false alarms too high for practical usage
- ★ Performance degradation with increasing number of attack types
- ★ Unable to distinguish similar but different attacks (U2R vs R2L, types of DDoS, ...)
- ★ Trade-off between speed and accuracy

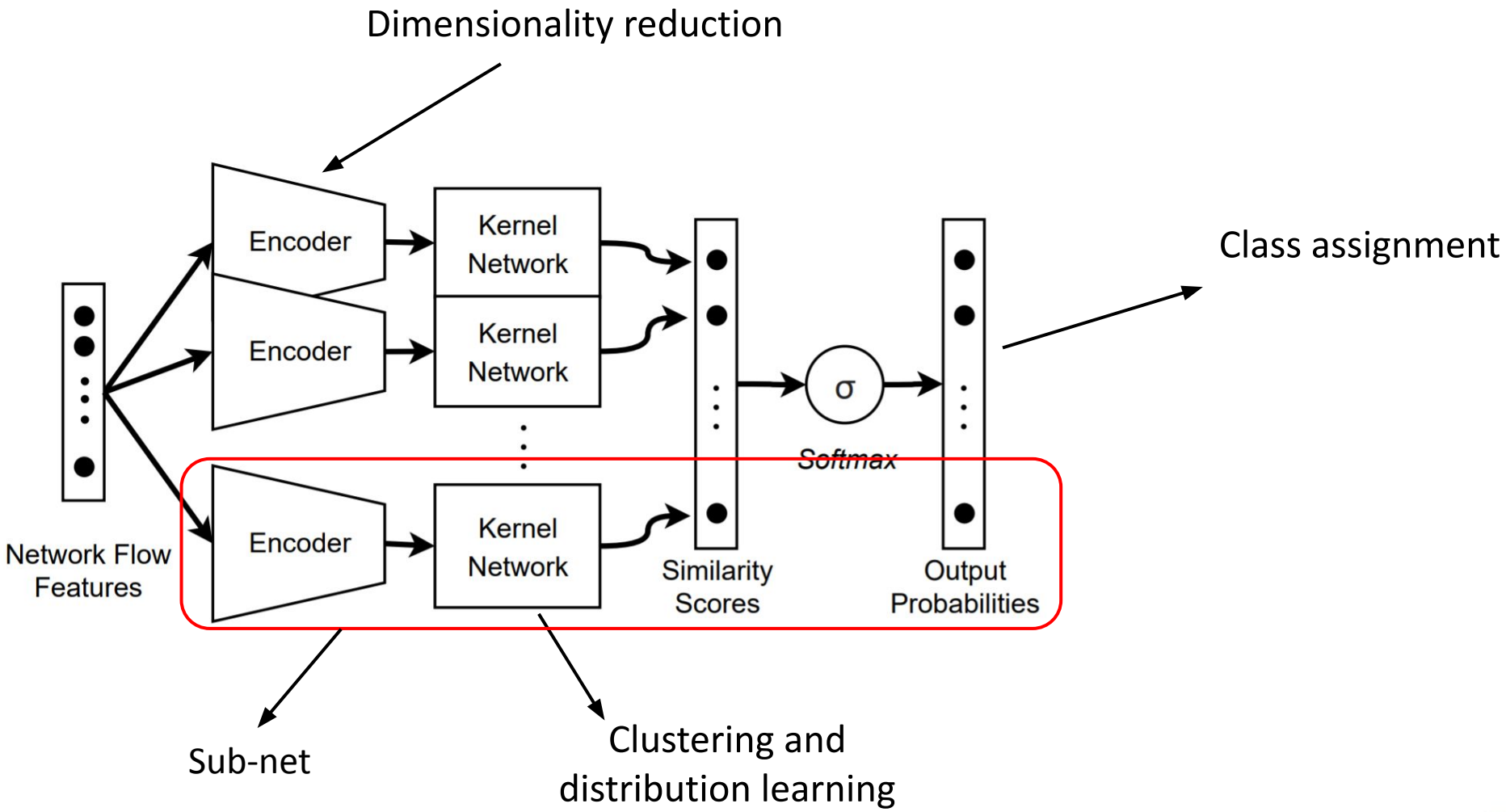
- Threat models: *Attacker inside/outside the LAN*

3 Proposed Solution

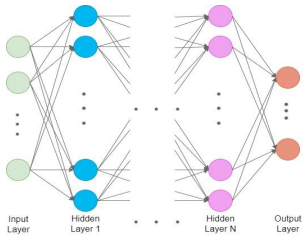


ACID Architecture: Adaptive Clustering-based Intrusion Detection

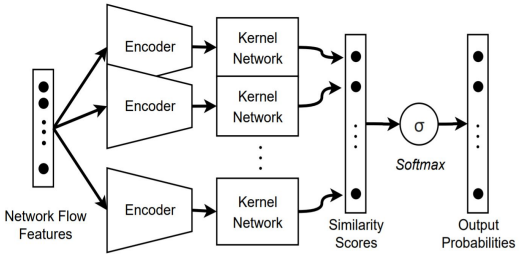
4 Solution | Adaptive Clustering network (AC-Net)



5 Solution | Adaptive Clustering network (AC-Net)

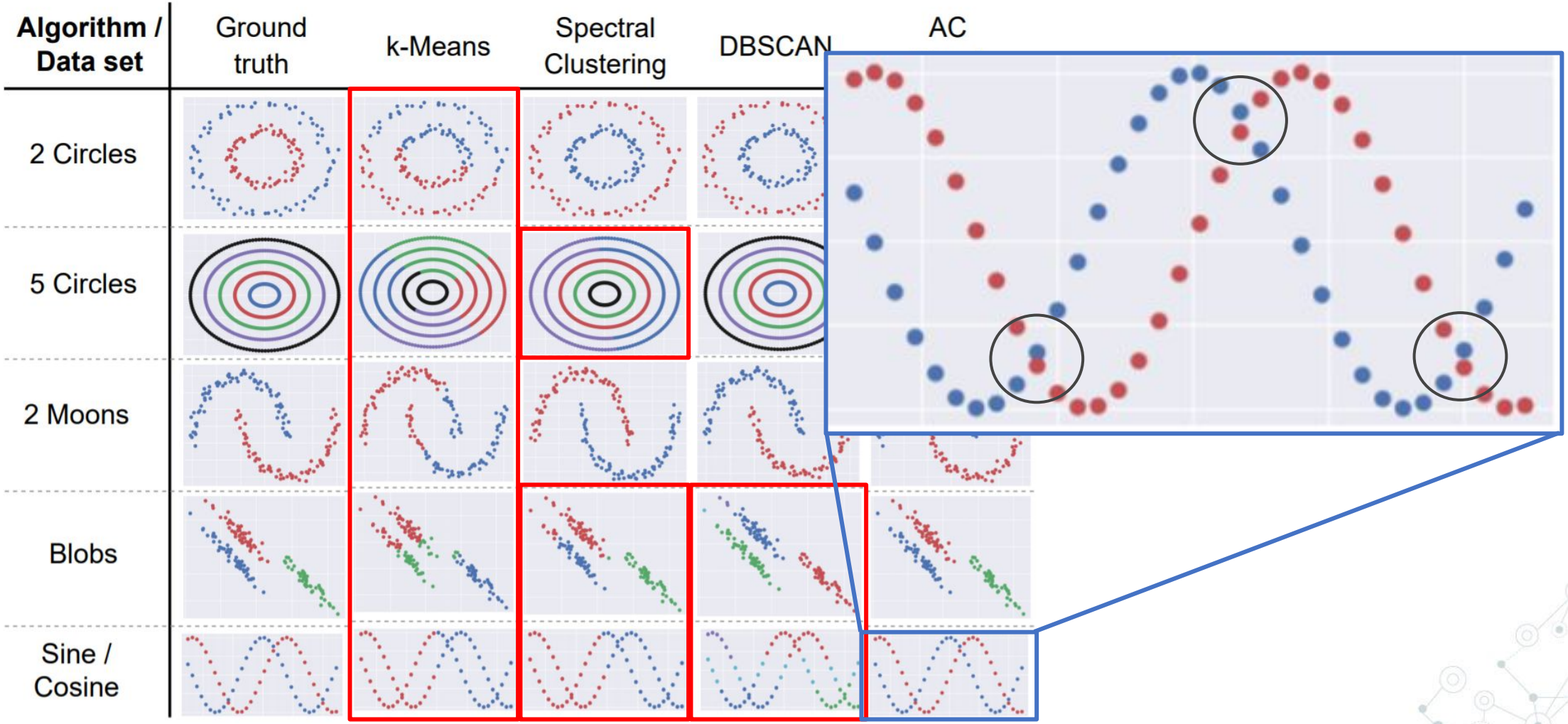


VS

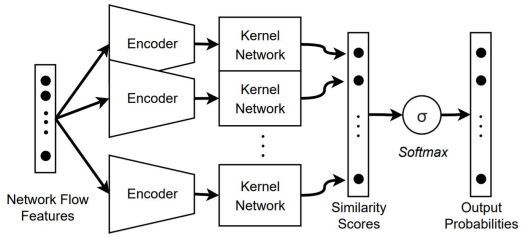


	Classical Neural Networks	***Adaptive Clustering Networks***
Scalability	Difficult	Easy
Parallelization	Data	Data + Sub-nets
Model Complexity	High (1 network = all tasks)	Low (1 sub-net = 1 task)
Architecture	Fixed (high risk of network saturation, conflicts in learned parameters)	Flexible (no network saturation, no conflict in learned parameters)
Sensitivity	Extreme (input features, unbalanced datasets, ...)	Marginal
Advantages	None	- Optimal class separation - Intrinsic support for continual learning - Built-in clustering mechanism

6 Results | Clustering with AC-Net



7 Results| AC-Net

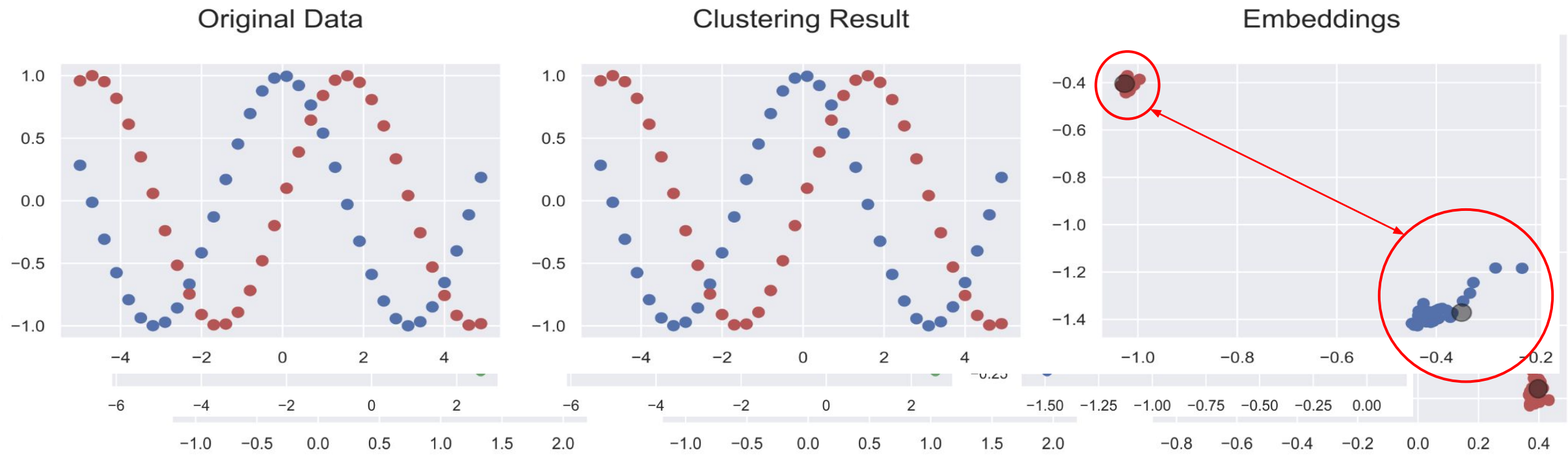


Embeddings

2 Circles

Blobs

Sine / Cosine



Results | Intrusion Detection

Binary classification
(Benchmark: ISCX-IDS 2012)

- **FAR:** False Alarm Rate

- *Classifier:* Random Forests
- Encoding dimension: 10
- Payload features: 50

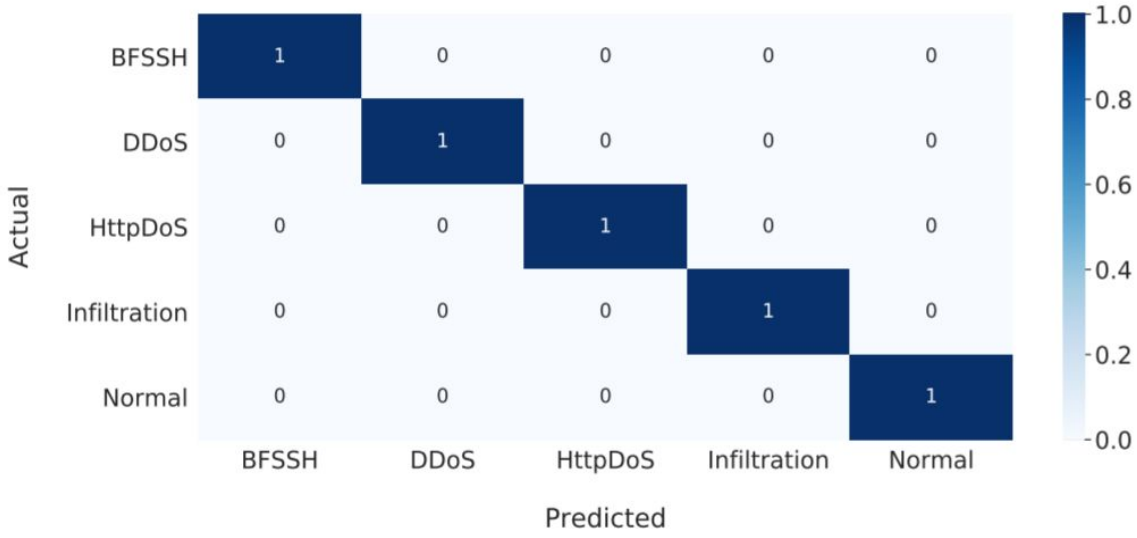
Approach	Payload-based Features	Accuracy (%)	FAR (%)	F_1 Score (%)
DAGMM	No	62.91	30.65	53.07
N-BaIoT	No	89.19	10.80	89.19
Deep NN	No	88.14	7.41	70.35
TR-IDS	Yes	98.88	1.12	98.87
ACID (ours)	No	99.78	0.23	99.44
ACID (ours)	**Yes**	**100.0**	**0.00**	**100.0**

Comparison of ACID with existing methods

9 Results | Intrusion Detection

Multi-label classification (ACID)

Metric / Dataset	Accuracy (%)	FAR (%)	F_1 (%)	Classes	Samples
KDD CUP'99	100.0	0.00	100.0	23	43,510
ISCX-IDS 2012	100.0	0.00	100.0	5	10,547
CSE-CIC-IDS 2018	100.0	0.00	100.0	15	144,772



Normalized confusion matrix for multi-label classification using ACID on the ISCX-IDS 2012 dataset.

Properties

Datasets:

- Time span: *20 years*
- Number of attack types: *40*
- Raw network traffic traces
- Train/Test split: 70/30
- Payload features: Yes
- Test set ≅ 0.2 Billion packets

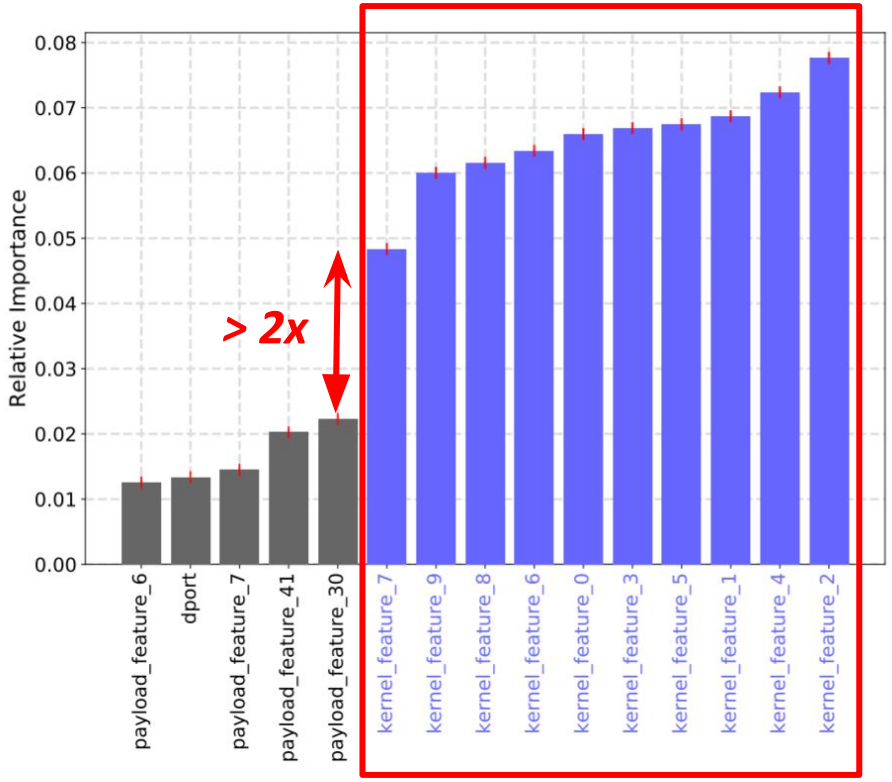
- *Classifier:* Random Forests
- Encoding dimension: 10
- Payload features: 50

10 Impact factors | ISCX-IDS 2012

- *Classifier:* Random Forests
- Encoding dimension: 10
- Payload features: 50

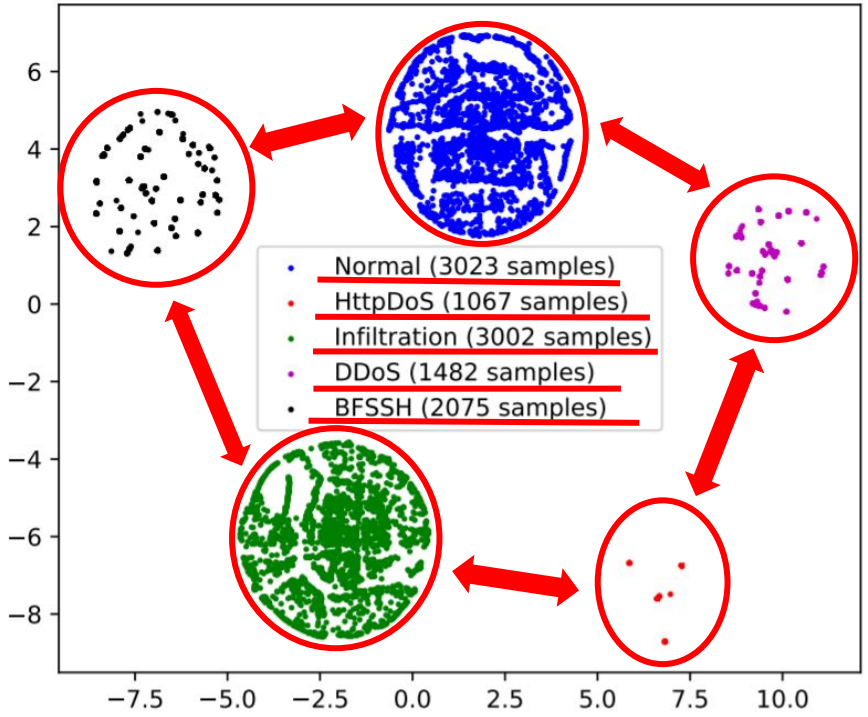
Feature ranking:

15 most important features in the classification process



t-SNE (from AC-Net's Embeddings)

t-SNE: A tool used to simplify the visual exploration of high-dimensional data points



11 Complexity Analysis

Speed Analysis (packets)

Payload features?	*Duration*	*Throughput*
No	0.78 us	1.3M pps
Yes	145 us	7K pps

Environmental setup

- 1 Virtual Machine
- 4 CPU cores @ 1.1GHz
- 4 GB RAM
- 50 GB Storage

✔ *Deployable on constrained devices*

> 100x speed up

Payload Features	Number of Parameters	Batch size	Model Complexity (MFLOP)	Execution Time (seconds)
No	789,855	1	1.49	0.08 ± 0.01
		128	191.68	0.10 ± 0.02
Yes	942,460	1	25.71	0.19 ± 0.04
		128	3291.43	18.59 ± 0.74

Questions ?

Read more ?

Alec F. Diallo, Paul Patras. *"Adaptive Clustering-based Malicious Traffic Classification at the Network Edge"* - **IEEE INFOCOM 2021.**

- PhD funded by arm

Source code available at:
github.com/Mobile-Intelligence-Lab/ACID

Contact: alec.frenn@ed.ac.uk