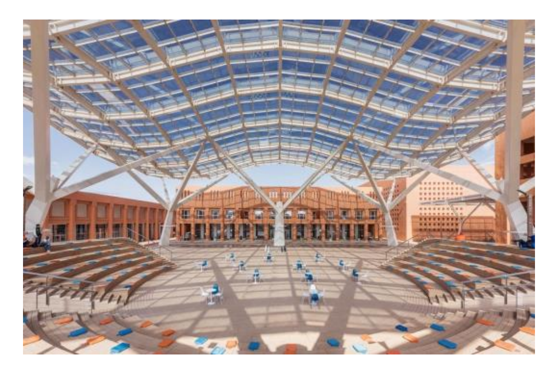
NF-NIDS: Normalizing Flows for Network Intrusion Detection Systems

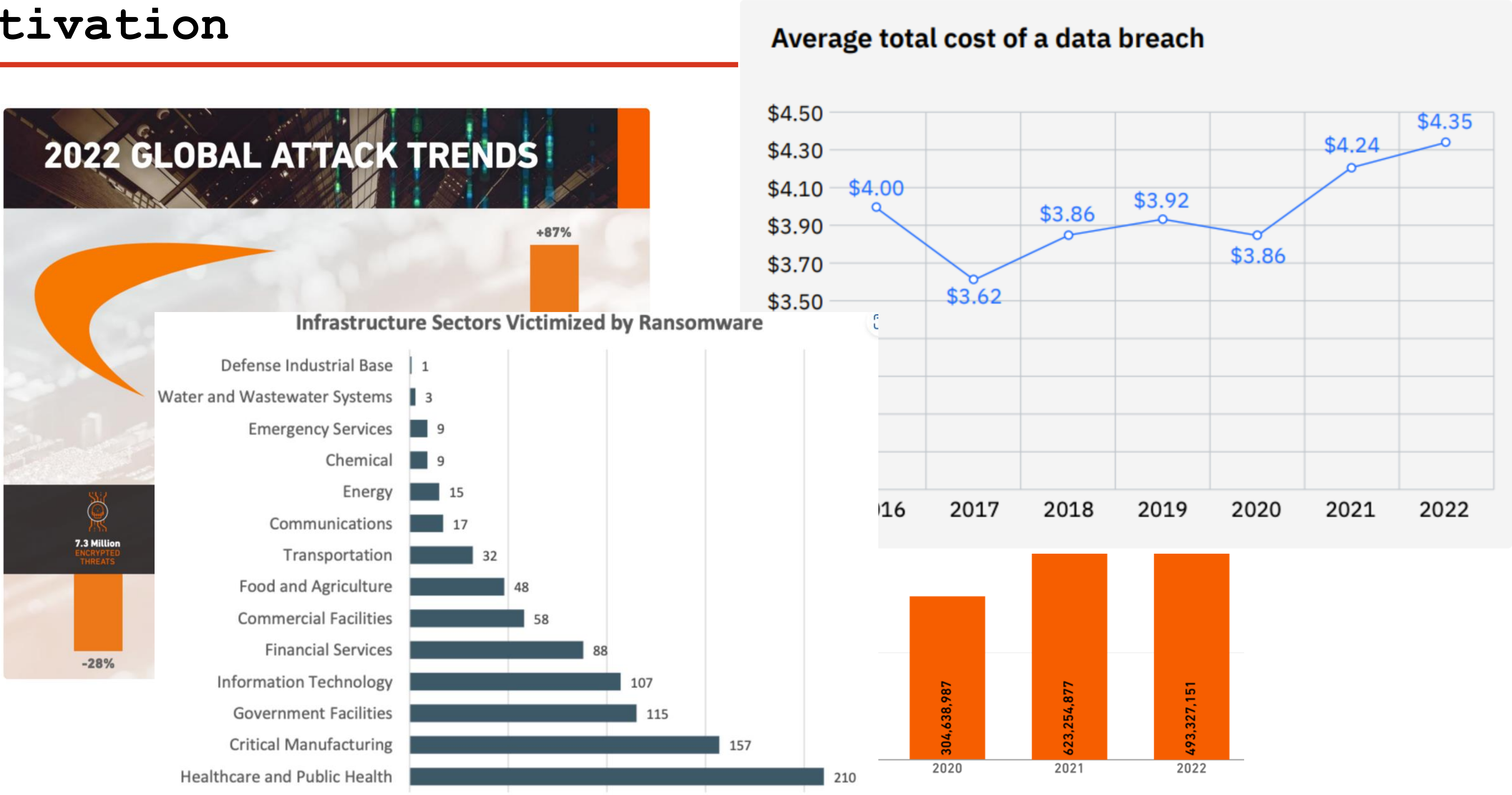
Meryem Janati Idrissi, Hamza Alami, Abdelhak Bouayad, Ismail Berrada

College of Computing
Mohammed VI Polytechnic University
October 27, 2023



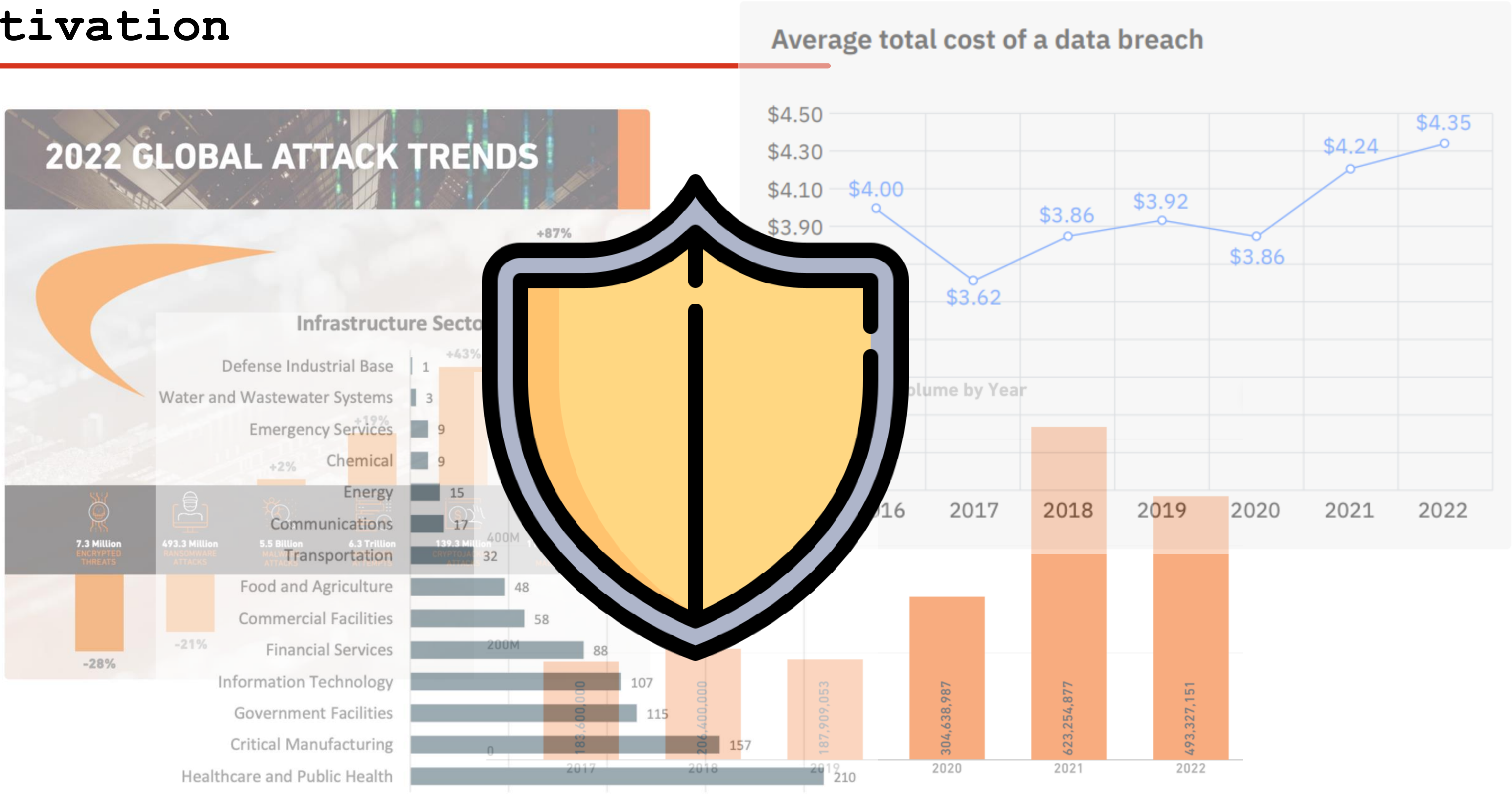
WINCOM'23

Motivation



Sources:Statista, CISCO, IBM and the Ponemon Institute, FBI IC3, ConnectWise Norton, Astra …

Motivation



Sources:Statista, CISCO, IBM and the Ponemon Institute, FBI IC3, ConnectWise Norton, Astra …

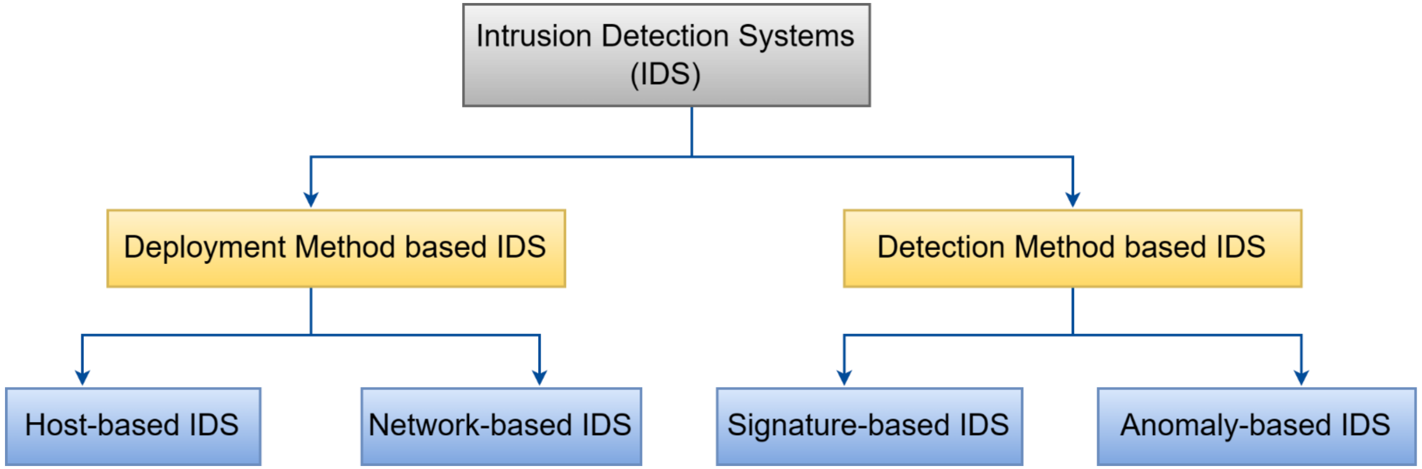
Intrusion Detection Systems (IDS)

- Defined as the tools, methods and technologies in computer and network security to detect and respond to unauthorized access.
- IDS can detect and deal with insider attacks, as well as external attacks.
- When it identifies such activity, it generates alerts or takes predefined actions to mitigate the threat.

IDS Taxonomy

- **Deployment Methods**
  - Host-based IDS: monitors the activities of a single host.
  - Network-based IDS: monitors the activities of all devices connected to a network.

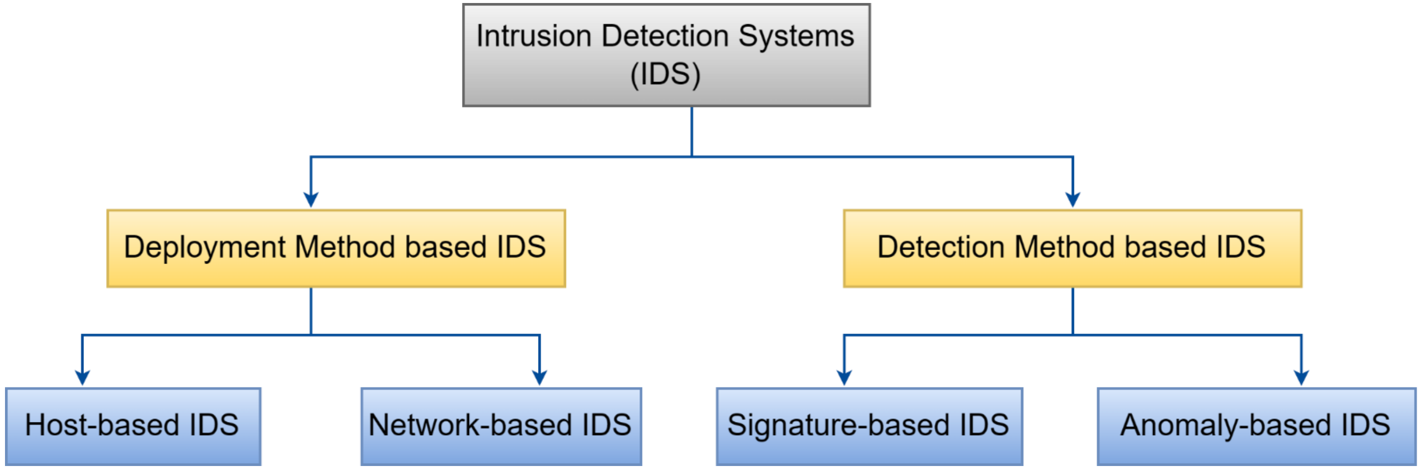
- **Detection Methods**
  - Signature-based IDS: relies on a database of known signatures to identify attack attempts.
  - Anomaly-based IDS: identifies possible deviations between the current event and the pre-established baseline of normal traffic profile.


Intrusion Detection Systems (IDS)
Deployment Method based IDS
Detection Method based IDS
Host-based IDS
Network-based IDS
Signature-based IDS
Anomaly-based IDS

IDS Taxonomy

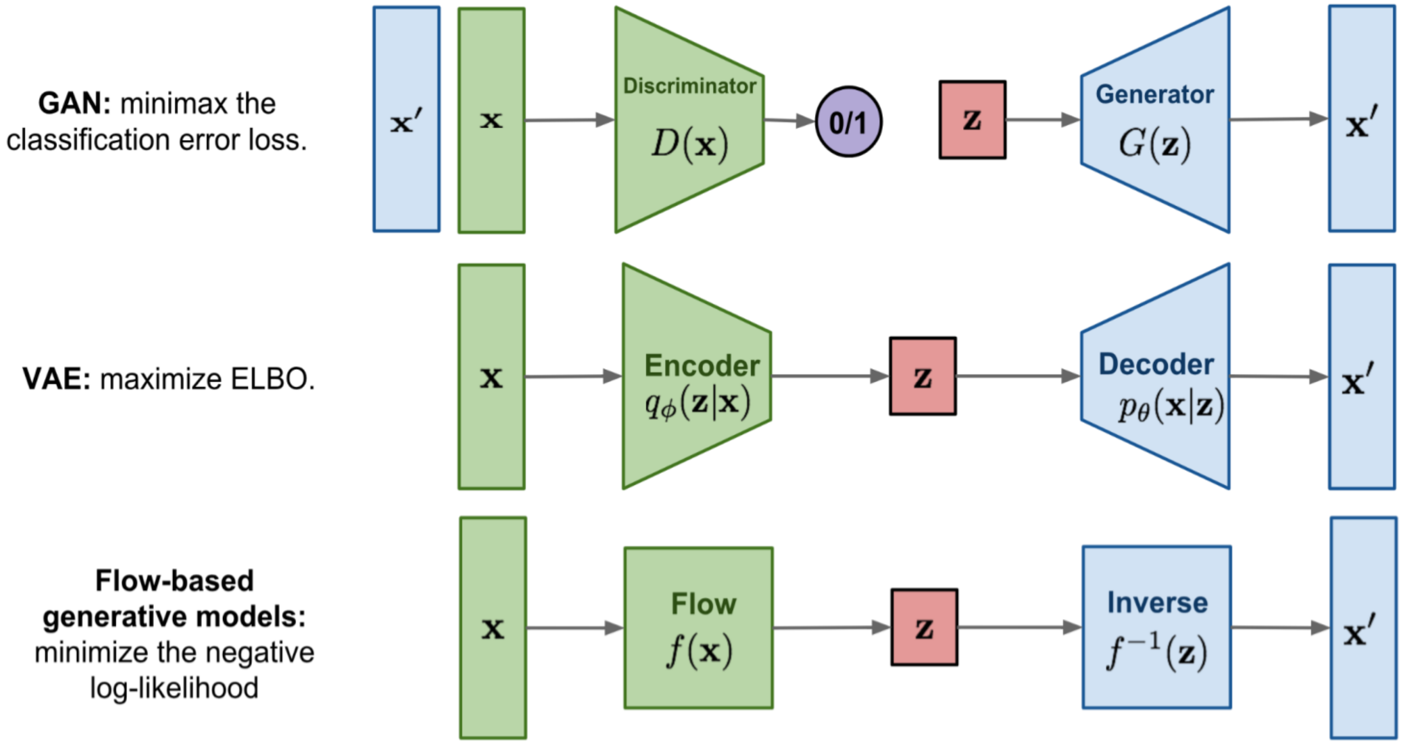
- **Deployment Methods**
  - Host-based IDS: monitors the activities of a single host.
  - Network-based IDS: monitors the activities of all devices connected to a network

- **Detection Methods**
  - Signature-based IDS: relies on a database of known signatures to identify attack attempts.
  - Anomaly-based IDS: identifies possible deviations between the current event and the pre-established baseline of normal traffic profile.

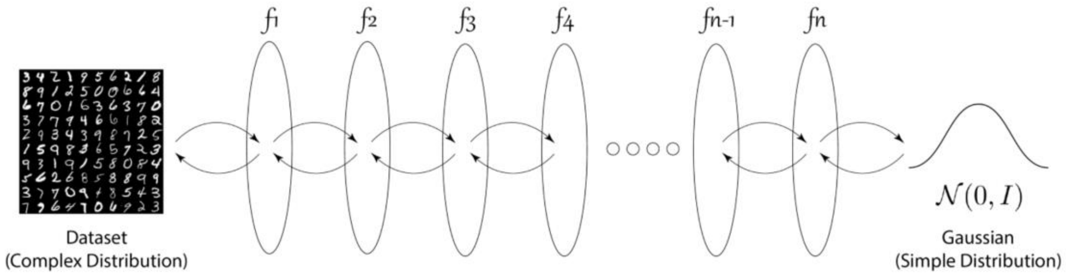

Intrusion Detection Systems (IDS)
Deployment Method based IDS
Detection Method based IDS
Host-based IDS
Network-based IDS
Signature-based IDS
Anomaly-based IDS

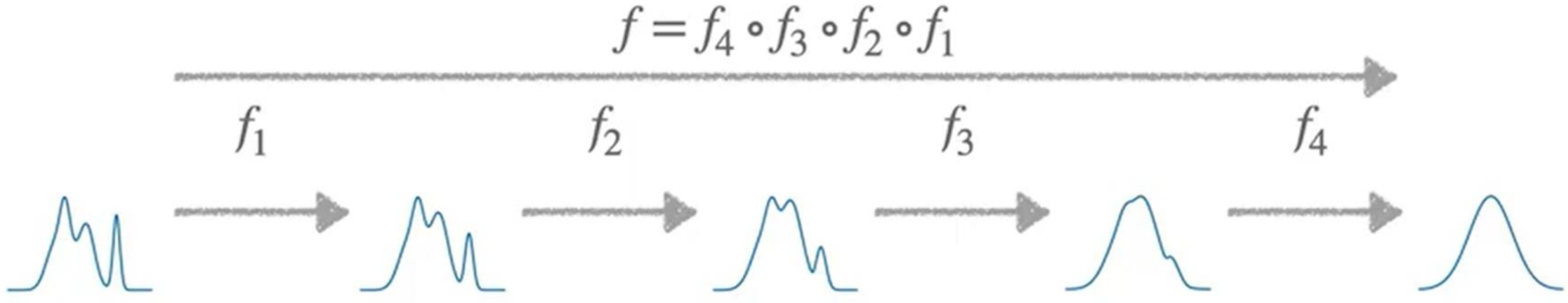
GAN & VAE for anomaly-based NIDS

GANs and VAEs have shown inspiring results for learning complex data distributions and having simple inference methods. However, Neither of them explicitly learns the probability density function of real data p(x) because it is really hard!


GAN: minimax the classification error loss.
x′
x
Discriminator
$D(\mathbf{x})$
0/1
z
Generator
$G(\mathbf{z})$
x′
VAE: maximize ELBO.
x
Encoder
$q_\phi(\mathbf{z}|\mathbf{x})$
z
Decoder
$p_\theta(\mathbf{x}|\mathbf{z})$
x′
Flow-based generative models: minimize the negative log-likelihood
x
Flow
$f(\mathbf{x})$
z
Inverse
$f^{-1}(\mathbf{z})$
x′

Normalizing Flows





A normalizing flow transforms a complex distribution into a simple one (Gaussian distribution) by applying a sequence of invertible transformation functions. Flowing through a chain of transformations, eventually we obtain a probability distribution of the final target variable.

Normalizing Flows

Given:

- A random variable of observed data $x \in X$.
- The latent variable $\mathrm{z} \in Z$.
- A simple base distribution $\mathrm{p_Z(z)}$ and the probability density $\mathrm{p_X}(x)$ over x.

Normalizing flows aim to learn a mapping function f, such that: $z = f(x)$ and $x = f^{-1}(\mathrm{z})$

$$f = (f_1 \circ f_2 \circ \ \dots \circ f_{\mathrm{k}})$$

- Density estimation can be performed using the forward mapping $\mathrm{X} \rightarrow \mathrm{Z}$
- Sampling can be achieved using the inverse mapping $\mathrm{Z} \rightarrow \mathrm{X}$.

Normalizing Flows

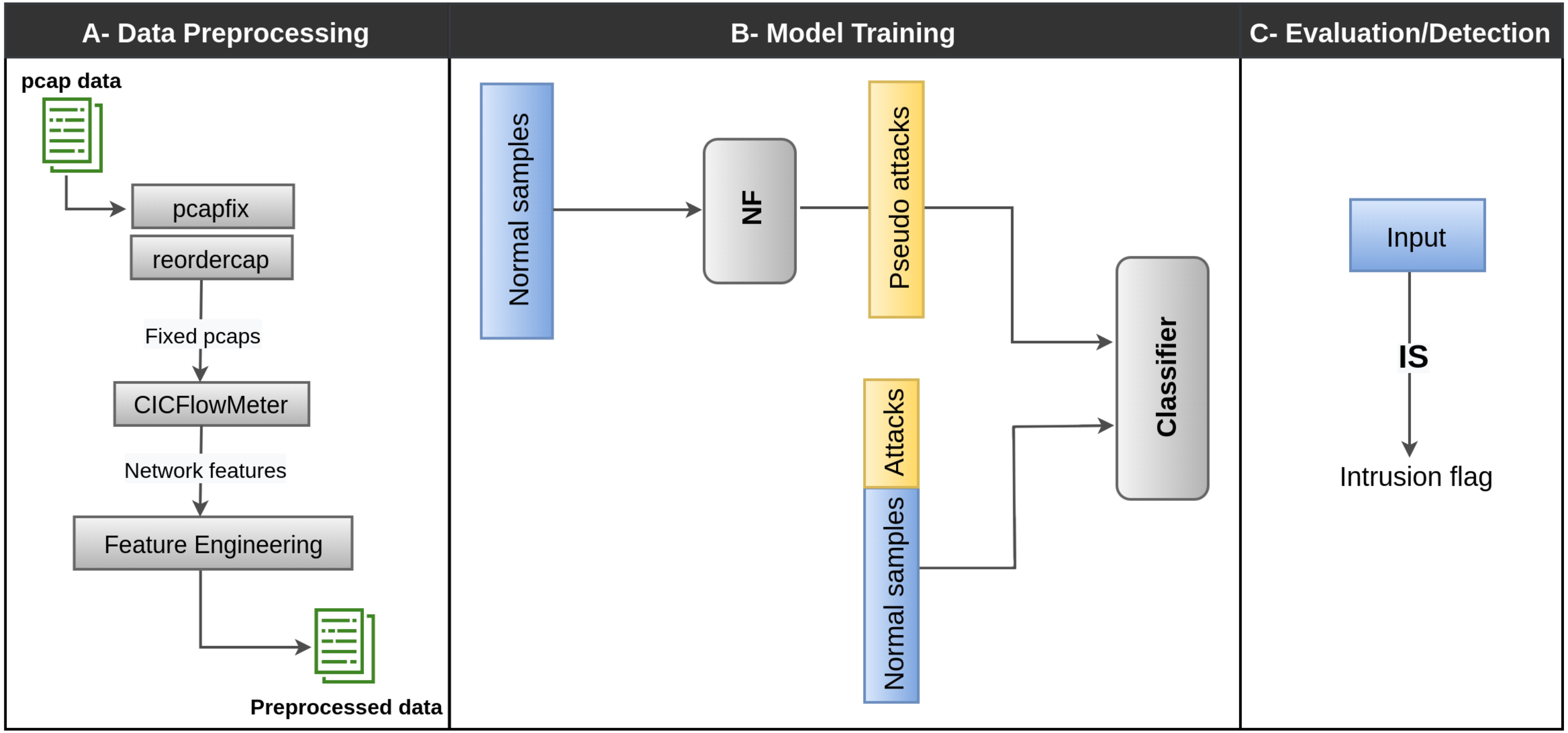
Change of variable formula:

$$p_X(x) = p_Z(z)\left|det\left(\frac{\partial z}{\partial x}\right)\right| = p_Z(f(x))\left|det\left(\frac{\partial f(x)}{\partial x}\right)\right|$$

Inverse Autoregressive Flow (IAF): the flow is designed such that the transformation from the latent space to the data space is autoregressive.

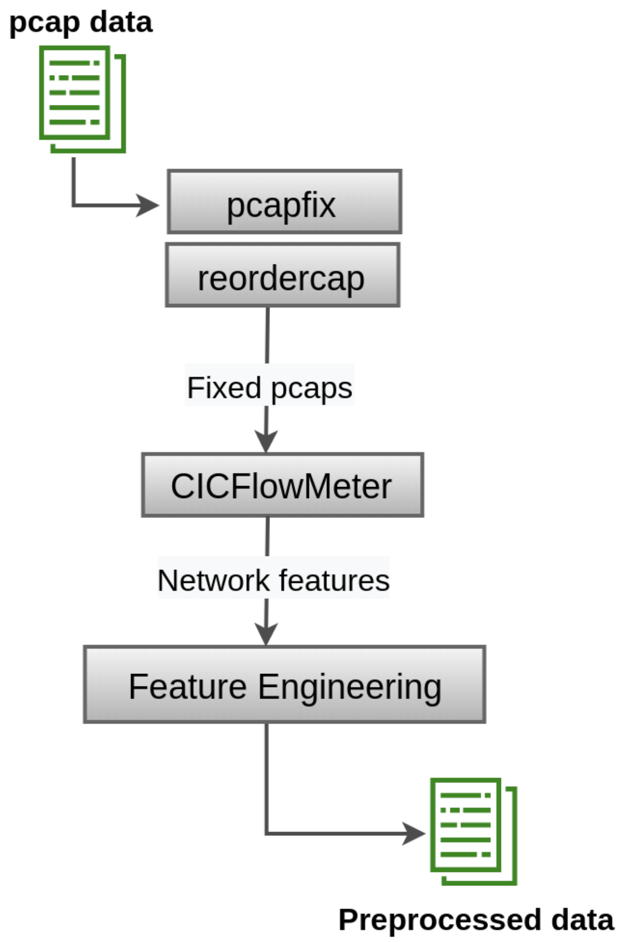
Neural Spline Flow (NSF): the transformation from z to x is constructed by composing a set of spline segments.

Our Method: NF-NIDS



NF-NIDS: Data preprocessing

- **Pcapfix:**
  - used to repair any possible damaged or corrupted pcap files.
- **Reordercap:**
  - ensures that the packets are ordered by timestamp.
- **CICFlowMeter:**
  - a widely used tool for feature extraction in network traffic analysis.
  - we adopt an improved version which addresses various issues related to labeling, flow construction, attack simulation …
  - a total of 87 features are extracted and saved into CSV files.
- **Feature engineering:**
  - Feature Conversion.
  - Feature Normalization (min-max scaling).



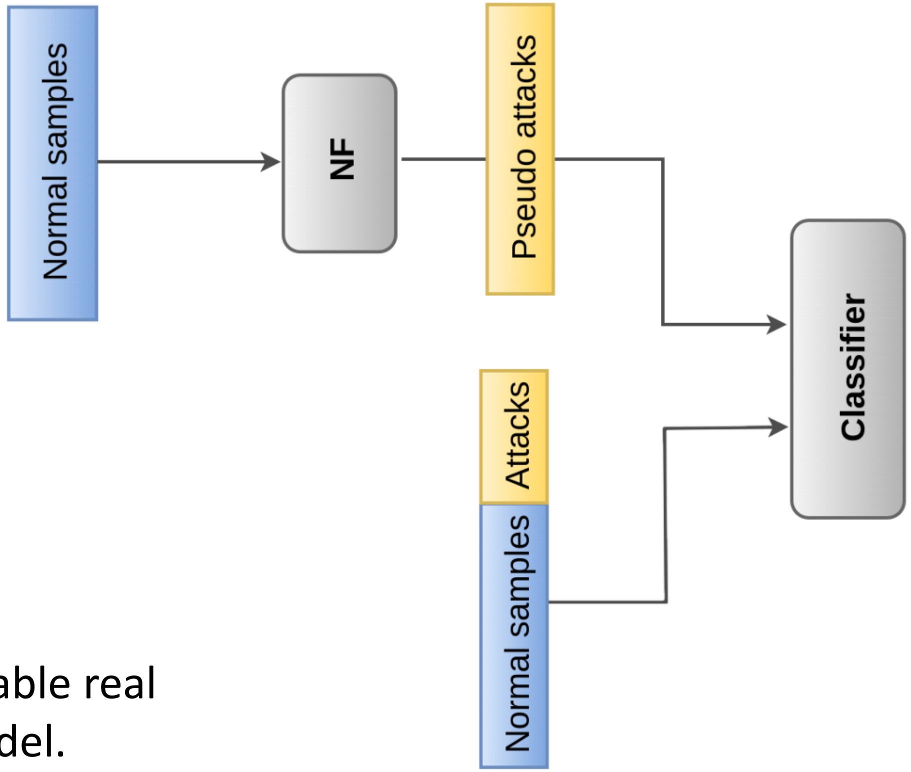
NF-NIDS: Model training

- **NF training:**
    - trained using only the set of normal network traffic data with the objective of maximizing the log-likelihood.

$$\mathcal{L} = -\frac{1}{|\mathcal{D}|}\sum_{x\in\mathcal{D}} log\, p_X(x)$$

    - it is used to sample pseudo-attack samples.

⚠ Assumption: only normal data is available

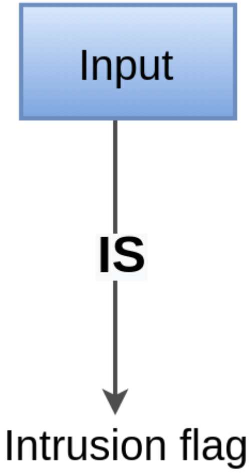


- **Classifier training:**
    - Trained using normal samples and a combination of the available real attacks and sampled pseudo-attacks generated by the NF model.

NF-NIDS: Detection phase

- **Classification:**
  - an Intrusion Score (IS) is computed for the input data, which, in our case, are the probabilities output by the classifier.
  - if the IS exceeds a certain threshold, the traffic is classified as an attack

$$x_{new} = \begin{cases} Normal, & \text{if IS} \leq \text{thr} \\ Anomaly, & \text{otherwise} \end{cases}$$



Experiments & Results

- **Datasets:**
  - **USTC-TFC2016:** The first part comprises 10 different categories of malware traffic. The second part of the dataset consists of normal traffic, representing 10 different genres (0.04% attacks used for training).
  - **CIC-IDS2017:** It comprises real-world network traffic data categorized as ”Benign” and ”Attacks. The dataset spans a five-day period, from Monday to Friday (0.08% attacks used for training).
  - **CSE-CIC-IDS2018:** is an enhanced version of the renowned CIC-IDS-2017 dataset (0.02% attacks for training).

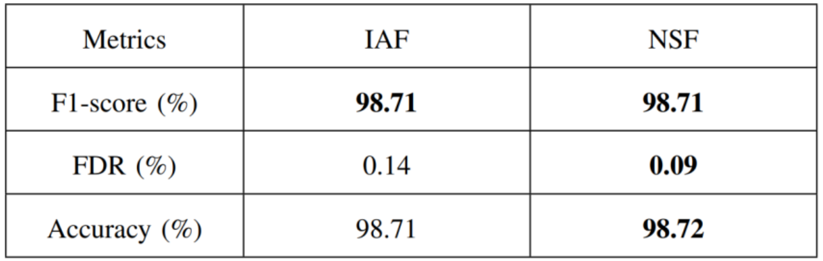
- **Model's architecture:**
  - IAF & NSF: 10 flows of neural networks as bijections, where each network is formed with two hidden layers.
  - Classifier: three fully-connected layer network.

- **Threshold selection:**
  - We empirically determine the threshold by testing the values in the range {0.85, 0.86, . . . , 0.99} and selecting the value that maximizes the F1-score in the testing set.

Experiments & Results

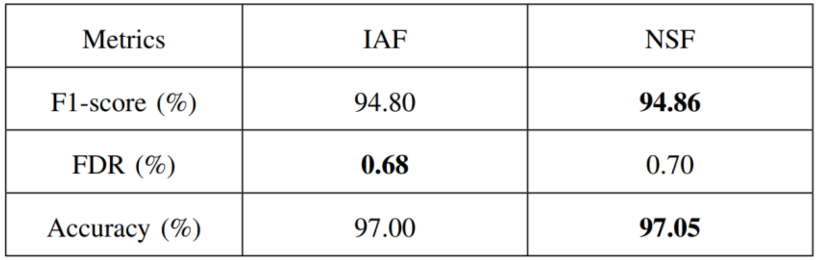
- The models can separate normal from attack samples with high accuracy and F1-score.
- FDR metric reports extremely low values, especially for the CSE-CIC-IDS2018.
- IAF and NSF models show comparable performance.
- IAF is remarkably faster than NSF.
- Incorporating the attack samples even when they are scarce, has been shown to improve the overall performance.

TABLE I: NF-NIDS testing results on USTC-TFC2016 dataset. The best results are presented in bold.



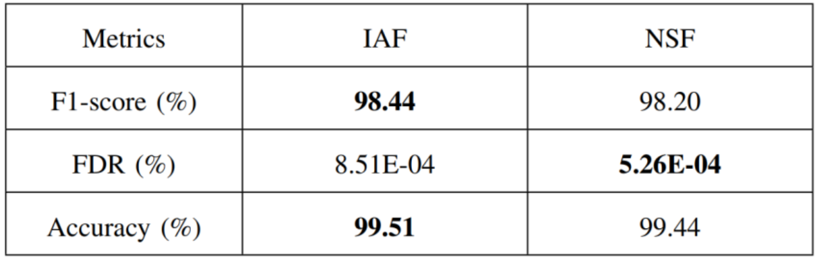
Metrics	IAF	NSF
F1-score (%)	**98.71**	**98.71**
FDR (%)	0.14	**0.09**
Accuracy (%)	98.71	**98.72**

TABLE II: NF-NIDS testing results on CIC-IDS2017.



Metrics	IAF	NSF
F1-score (%)	94.80	**94.86**
FDR (%)	**0.68**	0.70
Accuracy (%)	97.00	**97.05**

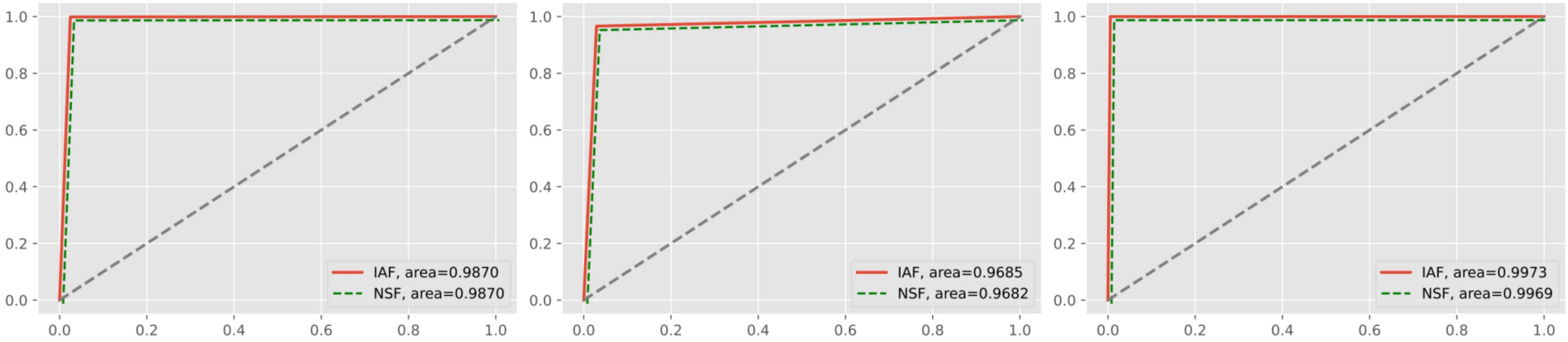
TABLE III: NF-NIDS testing results on CSE-CIC-IDS2018.



Metrics	IAF	NSF
F1-score (%)	**98.44**	98.20
FDR (%)	8.51E-04	**5.26E-04**
Accuracy (%)	**99.51**	99.44

Experiments & Results

- The ROC curves and AUC values further confirm the effectiveness of normalizing flows in network intrusion detection.
- For all datasets, the proposed method achieves the AUC of a minimum of 0.96%.



(a) USTC-TFC2016 (b) CIC-IDS2017 (c) CSE-CIC-IDS2018

Conclusion & perspectives

- We introduced NF-NIDS, a novel approach that leverages normalizing flows to accurately classify network traffic.
- Two main models, IAF and NSF, were employed to learn the exact distribution of normal behavior.
- We conducted empirical evaluations on multiple benchmark datasets, including USTC-TFC2016, CIC-IDS2017, and CSE-CIC-IDS2018, to demonstrate the effectiveness of our proposed method.
- Our models consistently achieved higher performance in terms of F1-score, accuracy, and false discovery rate.
- we aim to assess the effectiveness of other flow-based models, such as NICE, RealNVP, and Glow.

Conclusion & perspectives

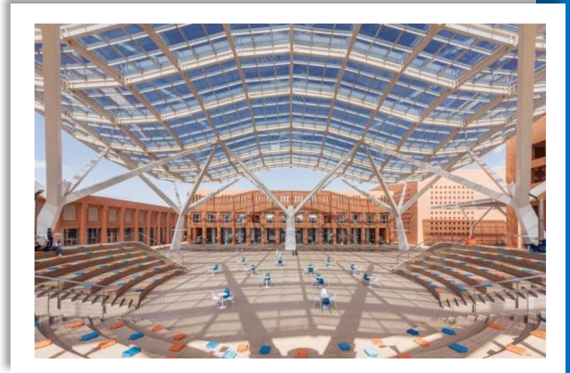
- We introduced NF-NIDS, a novel approach that leverages normalizing flows to accurately classify network traffic.
- Two main models, IAF and NSF, were employed to learn the exact distribution of normal behavior.
- We conducted empirical evaluations on multiple benchmark datasets, including USTC-TFC2016, CIC-IDS2017, and CSE-CIC-IDS2018, to demonstrate the effectiveness of our proposed method.
- Our models consistently achieved higher performance in terms of F1-score, accuracy, and false discovery rate.
- we aim to assess the effectiveness of other flow-based models, such as NICE, RealNVP, and Glow.

Thank you
for your attention!

NF-NIDS: Normalizing Flows for Network Intrusion Detection Systems

Meryem Janati Idrissi, Hamza Alami, Abdelhak Bouayad, Ismail Berrada

College of Computing
Mohammed VI Polytechnic University
October 27, 2023



WINCOM'23