MAGE-ID: A Multimodal Generative Framework for Intrusion Detection Systems

Mahdi Arab Loodaricheh, Mohammad Hossein Manshaei, Anita Raja

Department of Computer Science, Hunter College and The Graduate Center, City University of New York, NY, USA

Motivation

- Modern computer networks generate **high-dimensional, heterogeneous traffic**
- IDS (Intrusion Detection Systems) must handle evolving and sophisticated cyber threats
- Attack classes are often **underrepresented**
- Existing generative IDS methods are mostly **unimodal**
- Unimodal synthesis misses **cross-modal relationships**

Problem

- How can we generate realistic and balanced synthetic intrusion data?
- How can we preserve relationships across multiple representations?

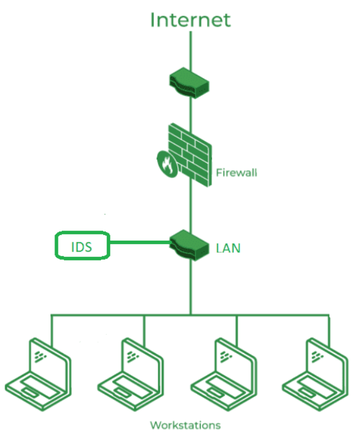
Hypothesis

Joint modeling of tabular data and transformed images can improve:

1. Realism, Diversity, Coverage
2. Downstream IDS Performance

Key Idea: Compact latent representations preserve key structure and enable stable multimodal generation.

IDS Datasets



CIC-IDS-2017
- 30,000 samples
- Benign vs. PortScan and DDoS
- 80:20 imbalance

NSL-KDD
- 30,000 samples
- Normal vs. Attack
- 80:20 imbalance

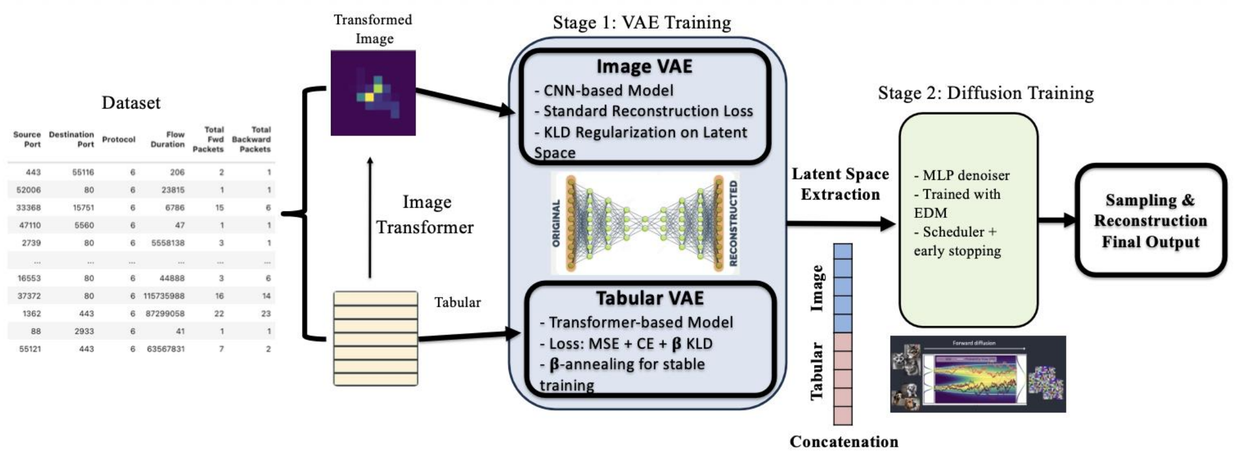
MAGE-ID Pipeline

- **Phase 1 – Data Transformation**
  - Start with tabular network-flow features
  - Transform features into images
  - Create paired tabular-image samples
- **Phase 2 – Representation Learning (Stage 1)**
  - Train a Transformer-based VAE (Variational Autoencoder) for tabular data
  - Train a CNN-based (Convolutional Neural Network) VAE for image data
  - Learn compact latent representations for both modalities
- **Phase 3 – Diffusion Modeling (Stage 2)**
  - Concatenate tabular and image latents
  - Model the joint latent space with an EDM-style denoiser
  - Generate coherent multimodal synthetic samples

$$\mathcal{L}_{\text{MAGE-ID}} = \text{MSE}(x_{\text{num}}, \hat{x}_{\text{num}}) + \text{CE}(x_{\text{cat}}, \hat{x}_{\text{cat}}) + \text{MSE}(x_{\text{img}}, \hat{x}_{\text{img}}) + \beta\, D_{\text{KL}}\big(q_\phi(z|x) \,\|\, \mathcal{N}(0, I)\big) + \mathbb{E}\big[w(\sigma)\|f_\theta(z + \sigma\varepsilon, \log\sigma) - z\|_2^2\big]\,.$$

- **Phase 4 – Reconstruction and Evaluation**
  - Decode latent samples into:
    - Synthetic tabular records
    - corresponding transformed images
  - Evaluation metrics:
    - PRDC (Precision, Recall, Density, Coverage) and Detectability
    - MLE (Machine Learning Efficacy)

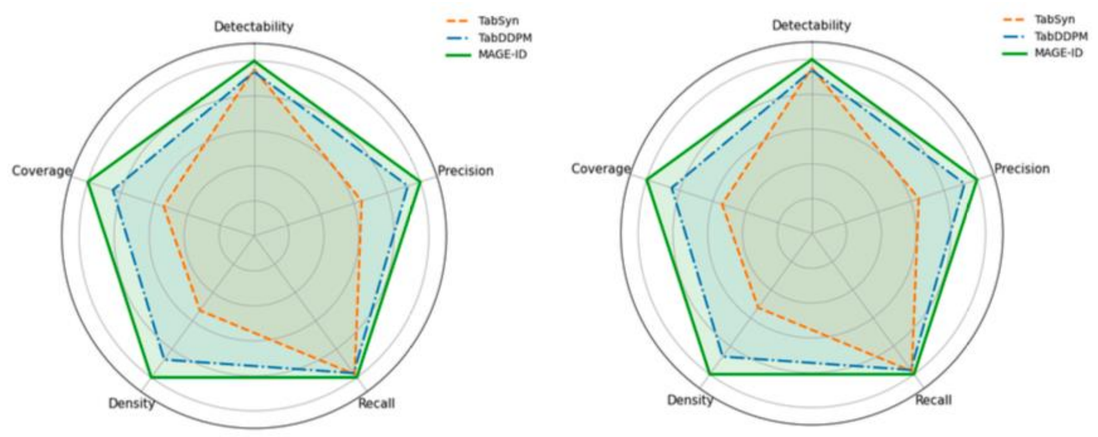
MAGE-ID Pipeline Overview



Results

TABLE I
PERFORMANCE COMPARISON OF GENERATIVE MODELS ON CIC-IDS-2017 AND NSL-KDD DATASETS

Model	Detectability	Precision	Recall	Density	Coverage	AUC (MLE)
CIC-IDS-2017 Dataset						
TabDDPM	0.9356	0.6993	0.9496	0.5210	0.5957	0.9937
TabSyn	0.9498	0.4894	0.9586	0.3140	0.3814	0.9976
MAGE-ID	**0.9974**	**0.7562**	**0.9795**	**0.5959**	**0.7017**	**0.9997**
NSL-KDD Dataset						
TabDDPM	0.9020	0.7541	0.9774	0.5801	0.6264	0.9683
TabSyn	0.9408	0.6553	0.9786	0.4877	0.6465	0.9809
MAGE-ID	**0.9543**	**0.7978**	**0.9808**	**0.6245**	**0.6882**	**0.9993**



- MAGE-ID outperformed TabSyn and TabDDPM on both datasets
- Best overall balance of fidelity and diversity
- Strong downstream utility for IDS training

Why MAGE-ID Works

- Transformed images preserve **feature correlations**
- Tabular and image modalities are **learned jointly**
- **Cross-representation** consistency improves realism
- **Multimodality** improves minority-class coverage

Future Work

- Add payload, PCAP, and system-log modalities
- Extend to **temporal modeling**
- Explore cross-dataset **transfer learning**
- Study broader cyber-defense applications

Acknowledgement

- Supported by the **Google Cyber NYC Institutional Research Program**
- Views expressed are those of the authors and do not necessarily reflect those of Google