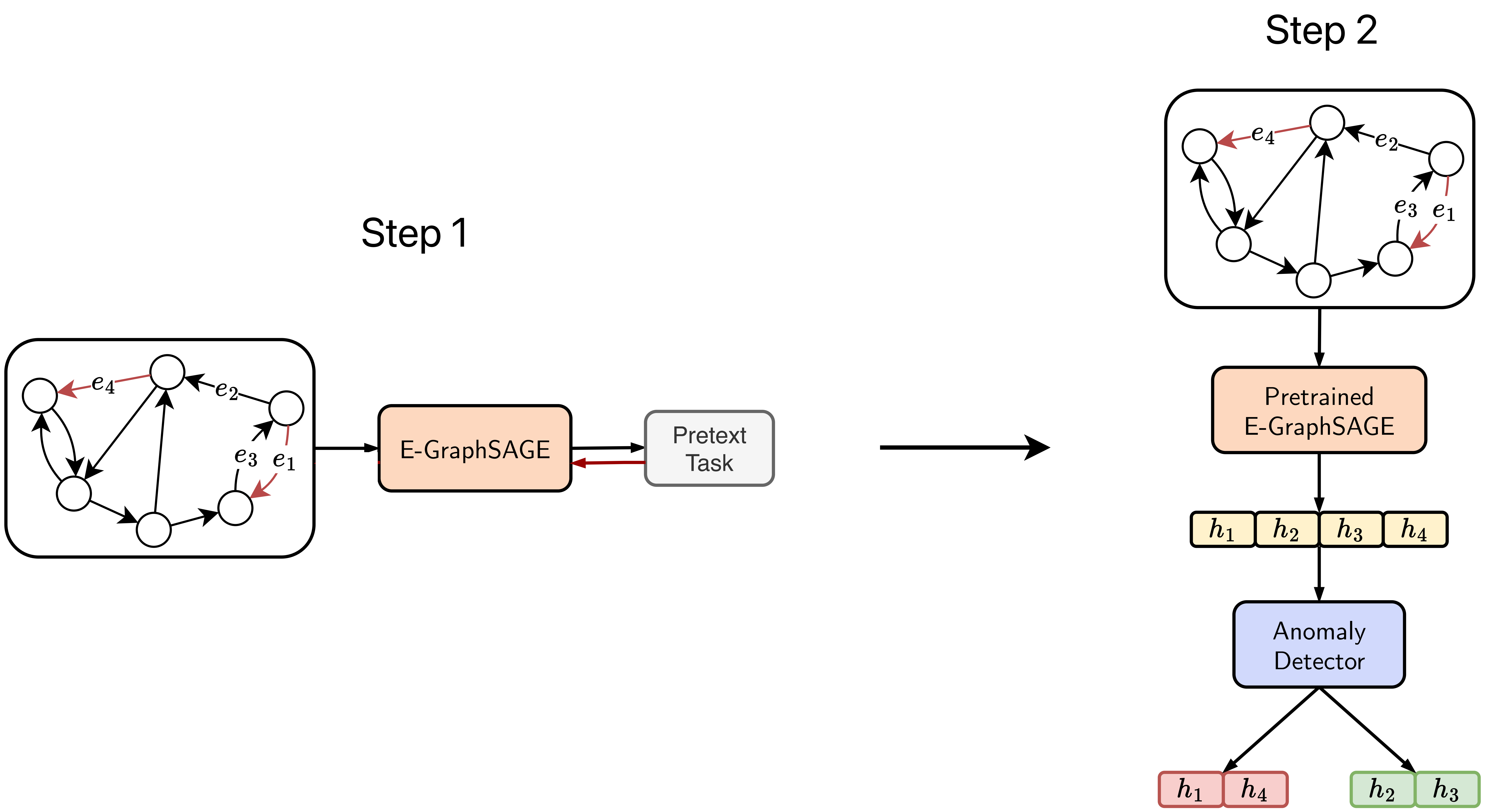
Self-Supervised Learning of Graph Representations for Network Intrusion Detection

Lorenzo Guerra, Thomas Chapuis, Guillaume Duc, Pavlo Mozharovskyi, Van-Tam Nguyen

Why is Intrusion Detection So Hard?

- **Supervised Models are a Dead End:** They require constant, expensive relabeling and are blind to zero-day attacks by design.
- **Self-Supervised Learning is the Only Path Forward:** It learns the network's normal behavior from massive, unlabeled data, but this requires highly expressive models.

The Problem with Existing Self-Supervised Models

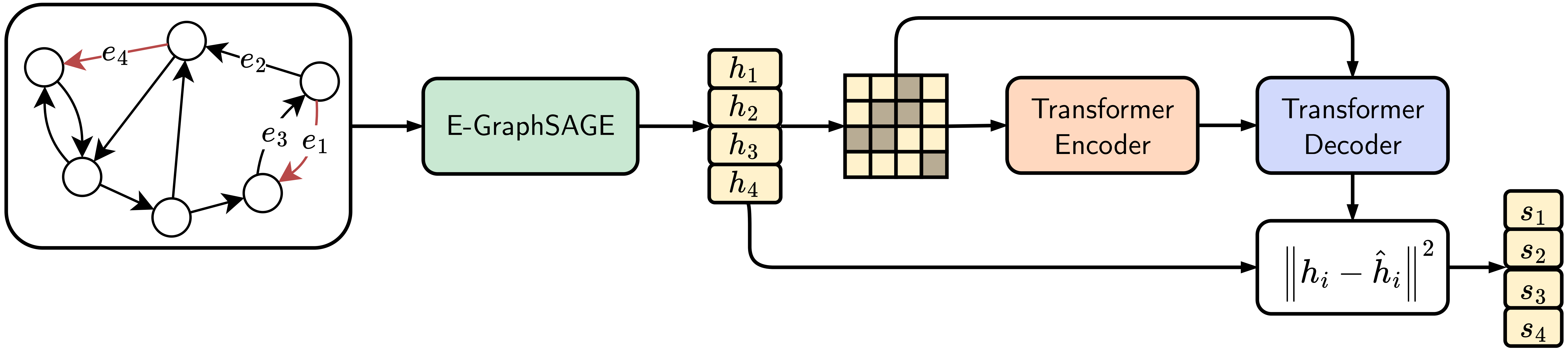


Our Solution: GraphIDS

1. **Unified End-to-End Framework:** Jointly trains a GNN and Transformer, forcing the GNN to learn embeddings directly optimized for anomaly detection.

2. **Local and Global Context:** E-GraphSAGE captures local topological patterns, while the Transformer's self-attention learns global co-occurrence patterns across the entire network.

3. **Simple & Effective Detection:** Anomaly score is simply the reconstruction error. No complex detectors or negative sampling needed



Key Results

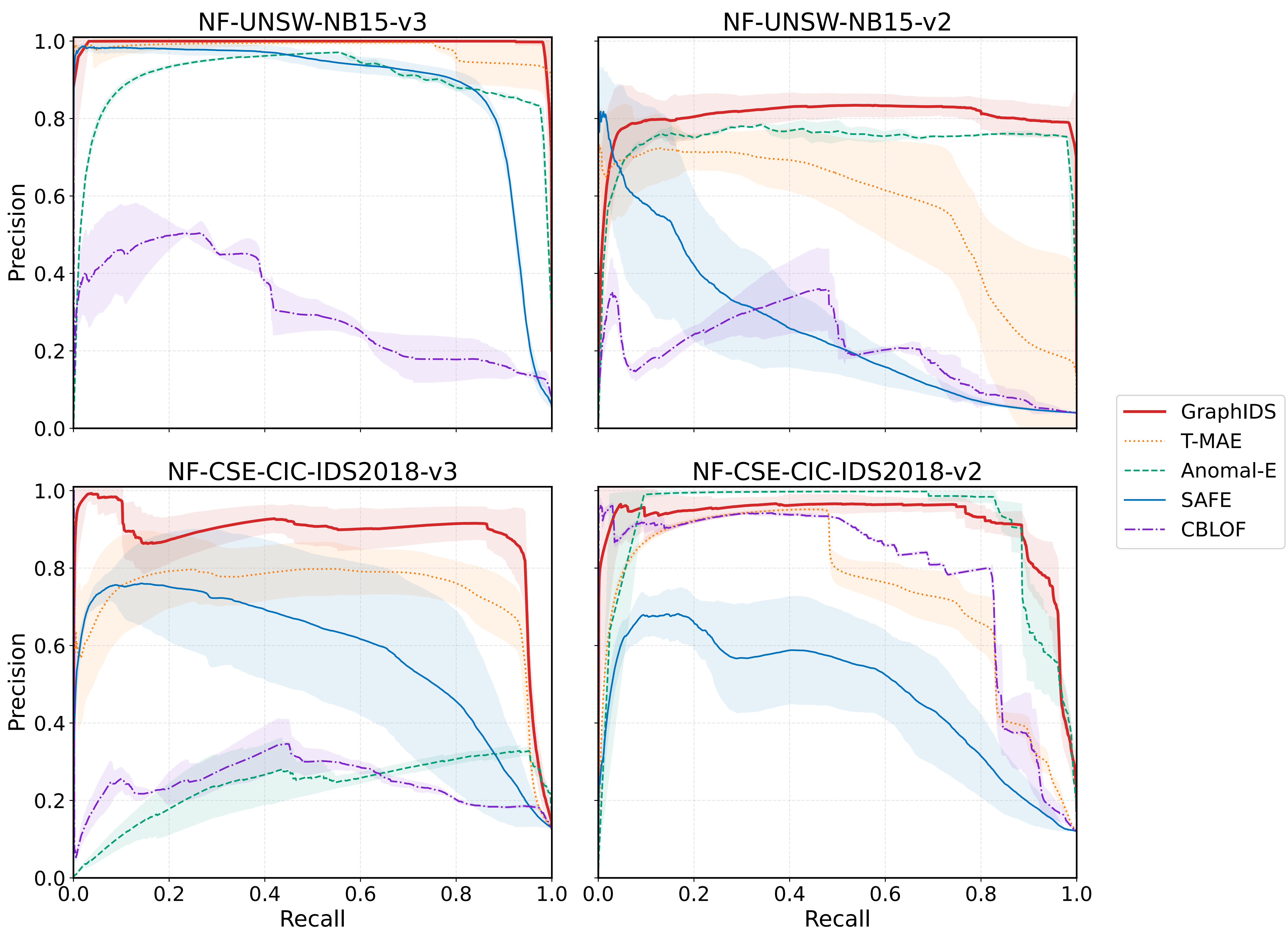


Figure 1. Precision-recall curves for all models on each dataset.

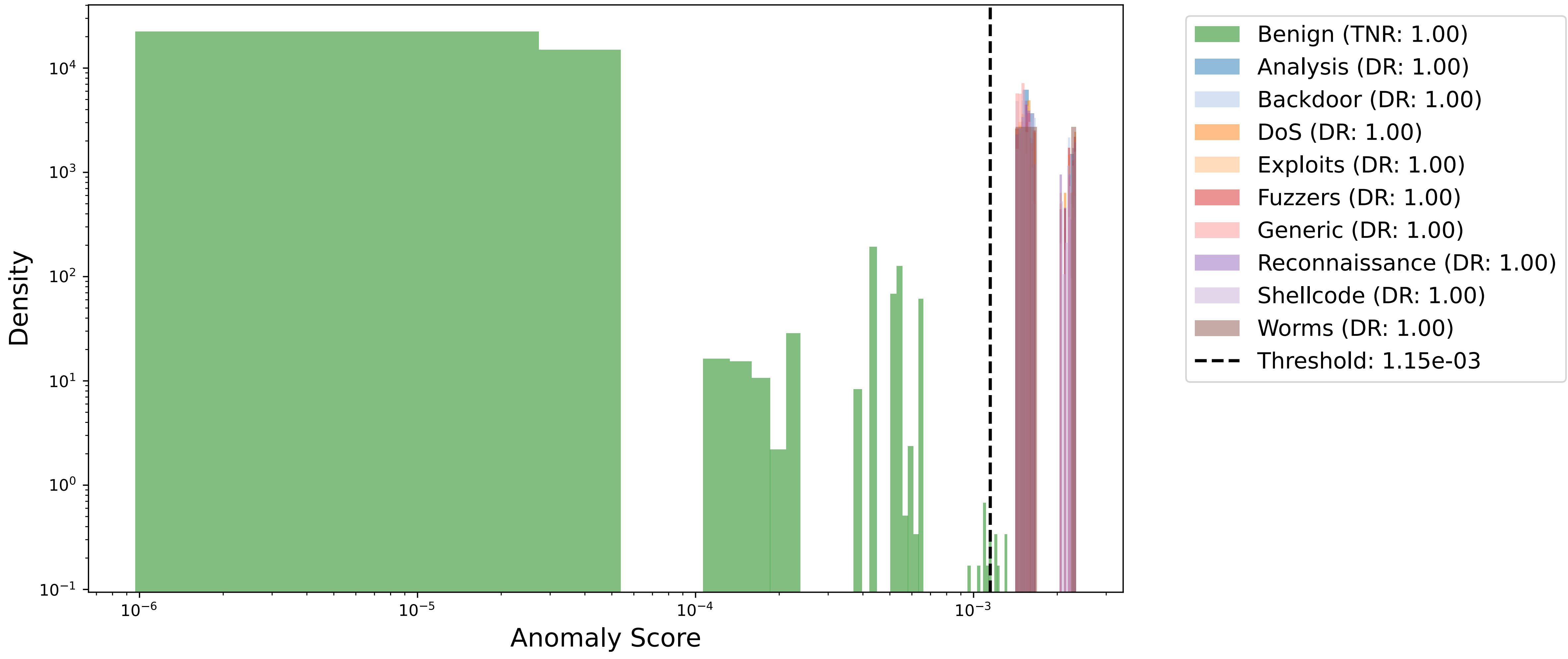


Figure 2. Anomaly score by attack type in NF-UNSW-NB15-v3.

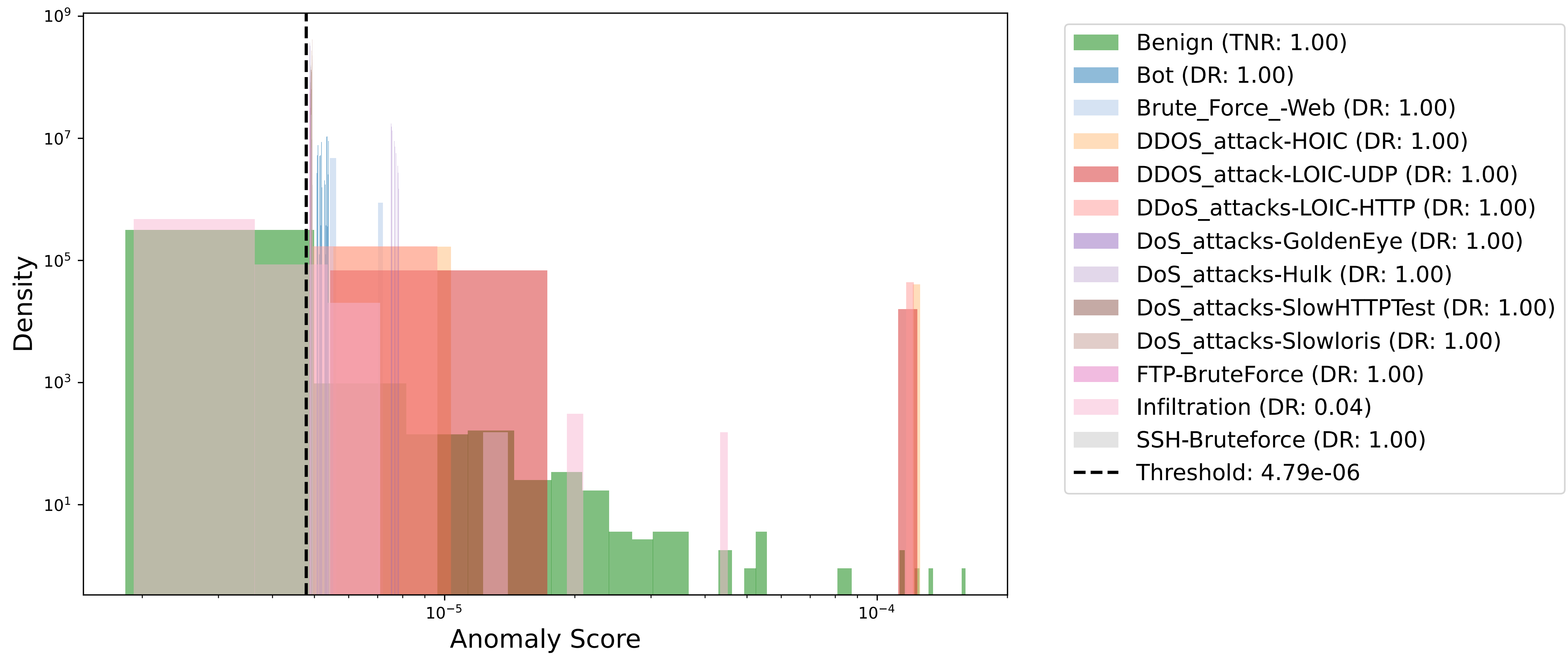


Figure 3. Anomaly score by attack type in NF-CSE-CIC-IDS2018-v3.

Conclusion & Takeaways

- **Introduced GraphIDS**: The first self-supervised framework to jointly train a GNN and a Transformer-based autoencoder for network intrusion detection.
- **Unified Representation**: The model learns by reconstructing graph-based flow embeddings, effectively unifying local topological context (from the GNN) with global co-occurrence patterns (from the Transformer).
- **State-of-the-Art Performance**: Achieves up to 99.98% PR-AUC, outperforming baselines by 5-25 percentage points, all without relying on labeled attack data or prior attack knowledge for training.