Utilizing Graph Neural Networks for Robust DDoS Attack Detection in Network Security

Kartikeya Sharma

Senior Associate Information Security Engineer at Equinix

What are Graph Neural Networks?

Graph Neural Networks are powerful AI tools that learn from connected data, helping us uncover hidden patterns in complex networks.

What are Graph Neural Networks?

Nodes (also known as vertices) represent entities or objects in a graph.

Edges represent the relationships or connections between nodes.

What are Graph Neural Networks?

GNNs learn rich node representations, called embeddings using Message Passing

What are Graph Neural Networks?

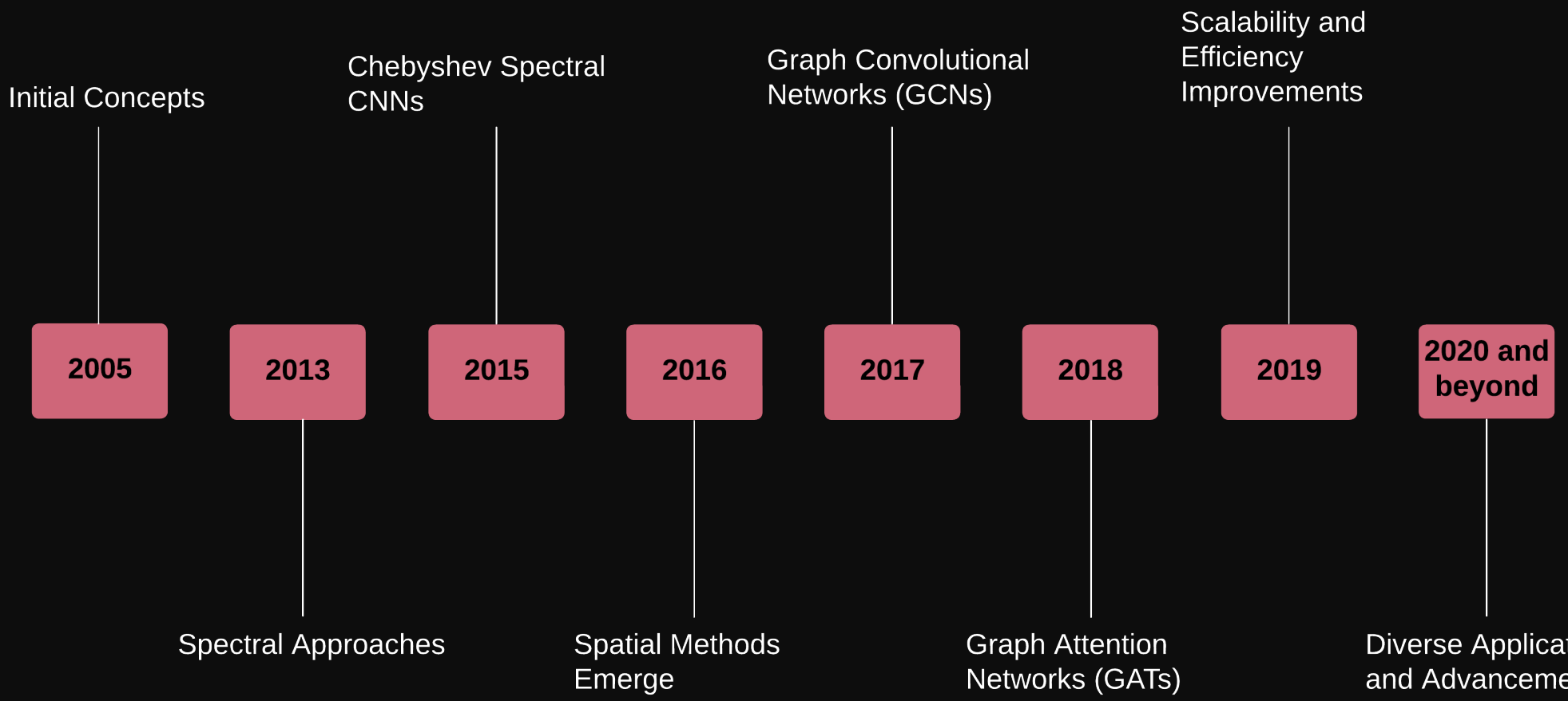
GNNs have found applications in various domains, including:

- ❑ Social network analysis
- ❑ Molecular property prediction
- ❑ Knowledge graph completion
- ❑ Recommender systems

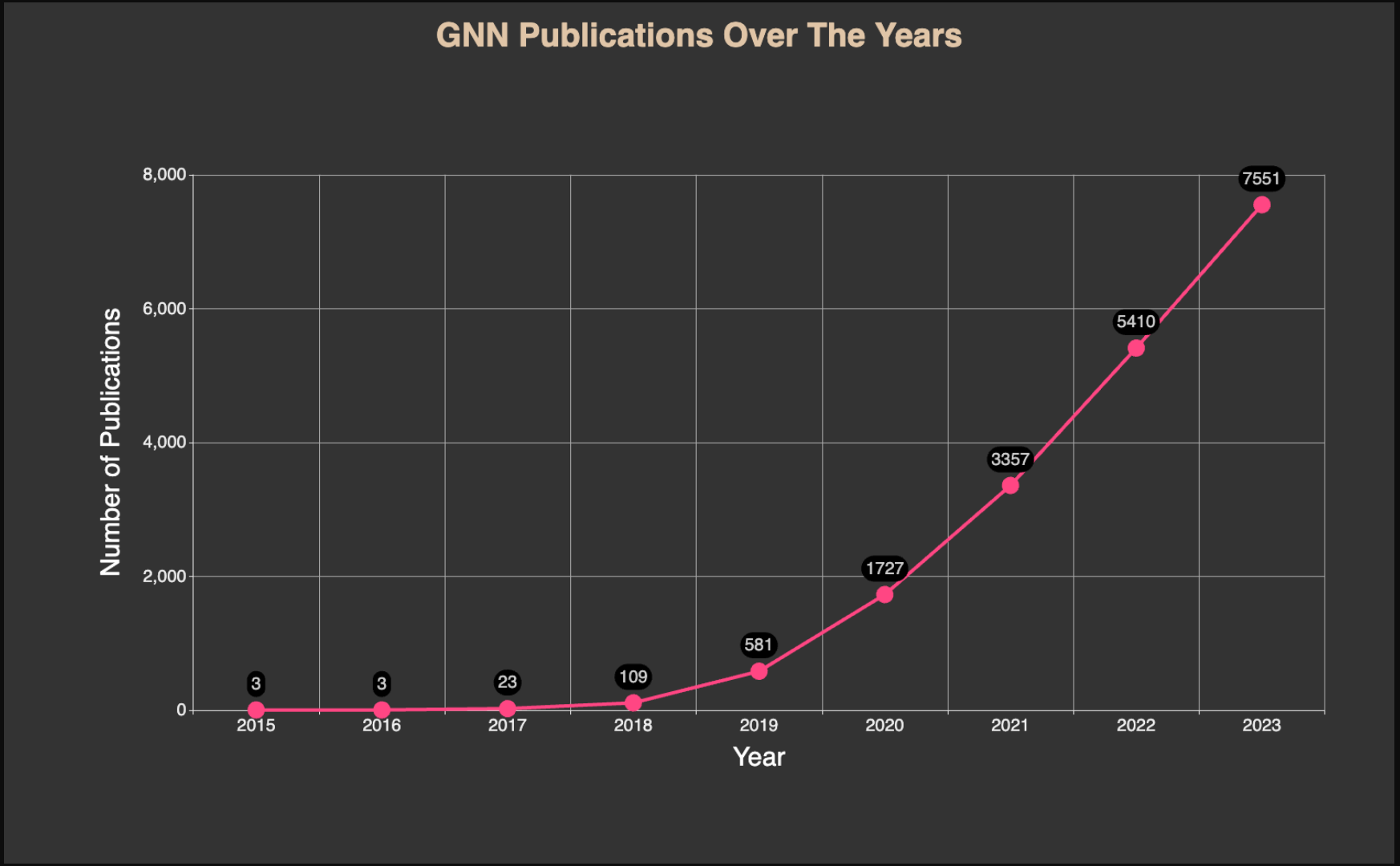
GNNs vs Traditional Neural Networks

Aspect	Graph Neural Networks	Traditional Neural Networks
Input Structure	**Graphs with variable size and connectivity**	**Fixed-size, grid-like input (e.g., images, sequences)**
Relationships	**Models and learns from relationships between entities**	**Assumes independence between input features**
Node-level Tasks	**Node classification, node regression, node clustering**	**Not applicable**
Edge-level Tasks	**Link prediction, edge classification**	**Not applicable**
Graph-level Tasks	**Graph classification, graph regression**	**Sample-level classification, regression**
Permutation Invariance	**Inherently permutation-invariant due to message passing**	**Requires explicit techniques (e.g., pooling) for permutation invariance**
Interpretability	**Can provide insights into important nodes, edges, and subgraphs**	**Often difficult to interpret learned features**

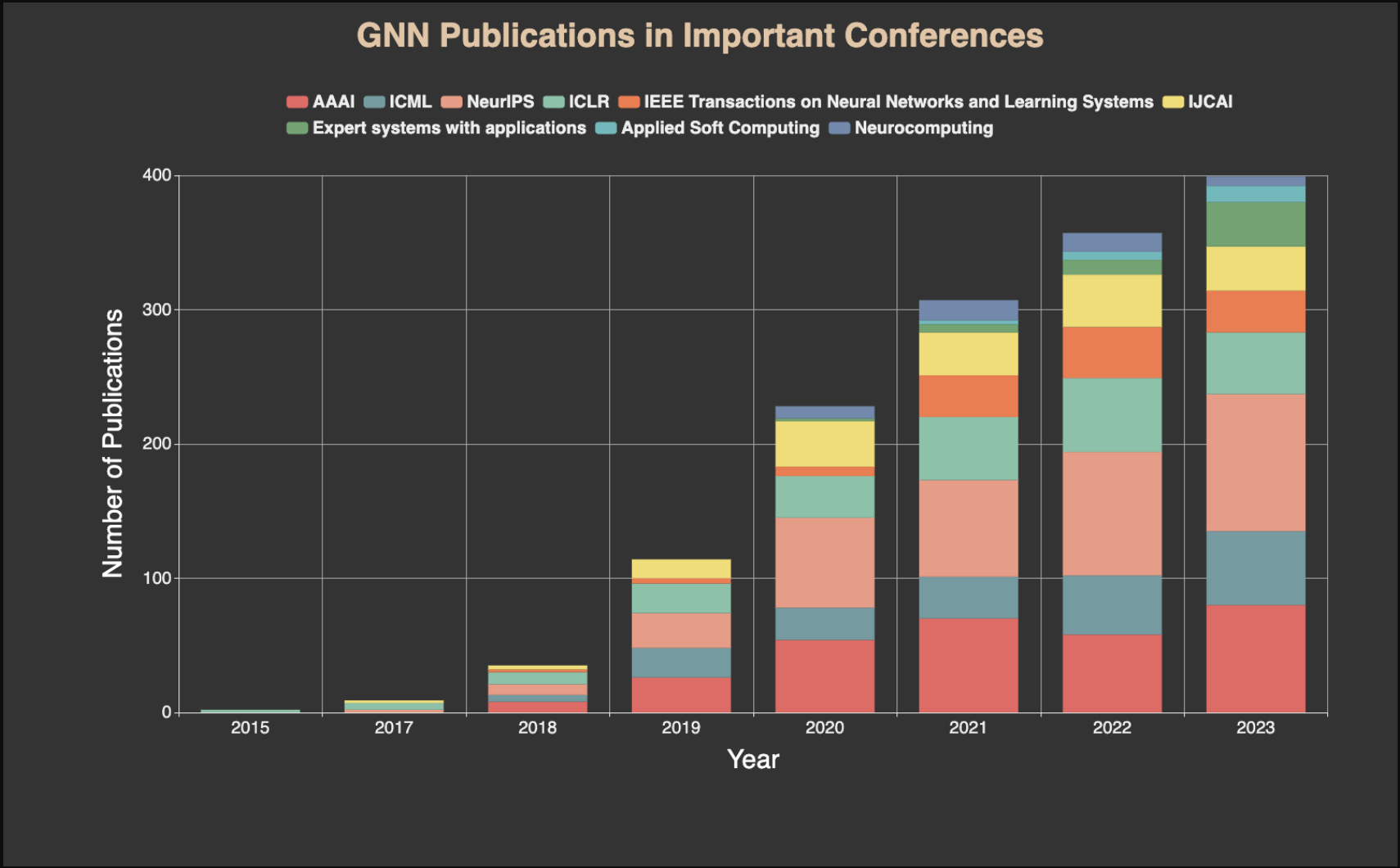
Milestones in GNN Evolution



Milestones in GNN Evolution



Milestones in GNN Evolution



What is a DDoS Attack?

A Distributed Denial of Service (DDoS) attack involves overwhelming a target—such as a server, website, or network—with a flood of internet traffic.

What is a DDoS Attack?

DDoS attacks can be categorized into three main types:

- ❑ Volume-based Attacks
- ❑ Protocol Attacks
- ❑ Application Layer Attacks

Traditional Approaches for DDoS Detection

- Filtering techniques
  - block traffic based on IP addresses, ports
- Statistical analysis
  - detect anomalies in traffic patterns, e.g. entropy, diversity
- Machine learning
  - k-Nearest Neighbors, Hidden Markov Models, Neural Networks

Traditional Approaches for DDoS Detection

Advantages of using traditional approaches:

- ❑ Simplicity and Low computational overhead
- ❑ Effectiveness against known attacks
- ❑ Interpretability

Traditional Approaches for DDoS Detection

Disadvantages of using traditional approaches:

- ❑ Limited adaptability
- ❑ Inability to model complex relationships
- ❑ High false positive rates
- ❑ Difficulty detecting low-volume attacks

The GNN Approach

- Represents the network as a graph
- Node features
  - IP address, port, and traffic statistics
- Edge features
  - Bandwidth and latency
- Learn node and edge embeddings and detect malicious activity by classifying nodes or entire graphs.

The GNN Approach

Advantages of using GNN approach:

- ❑ Automated feature learning
- ❑ Modeling complex relationships
- ❑ Generalization to unseen data

The GNN Approach

Disadvantages of using GNN approach:

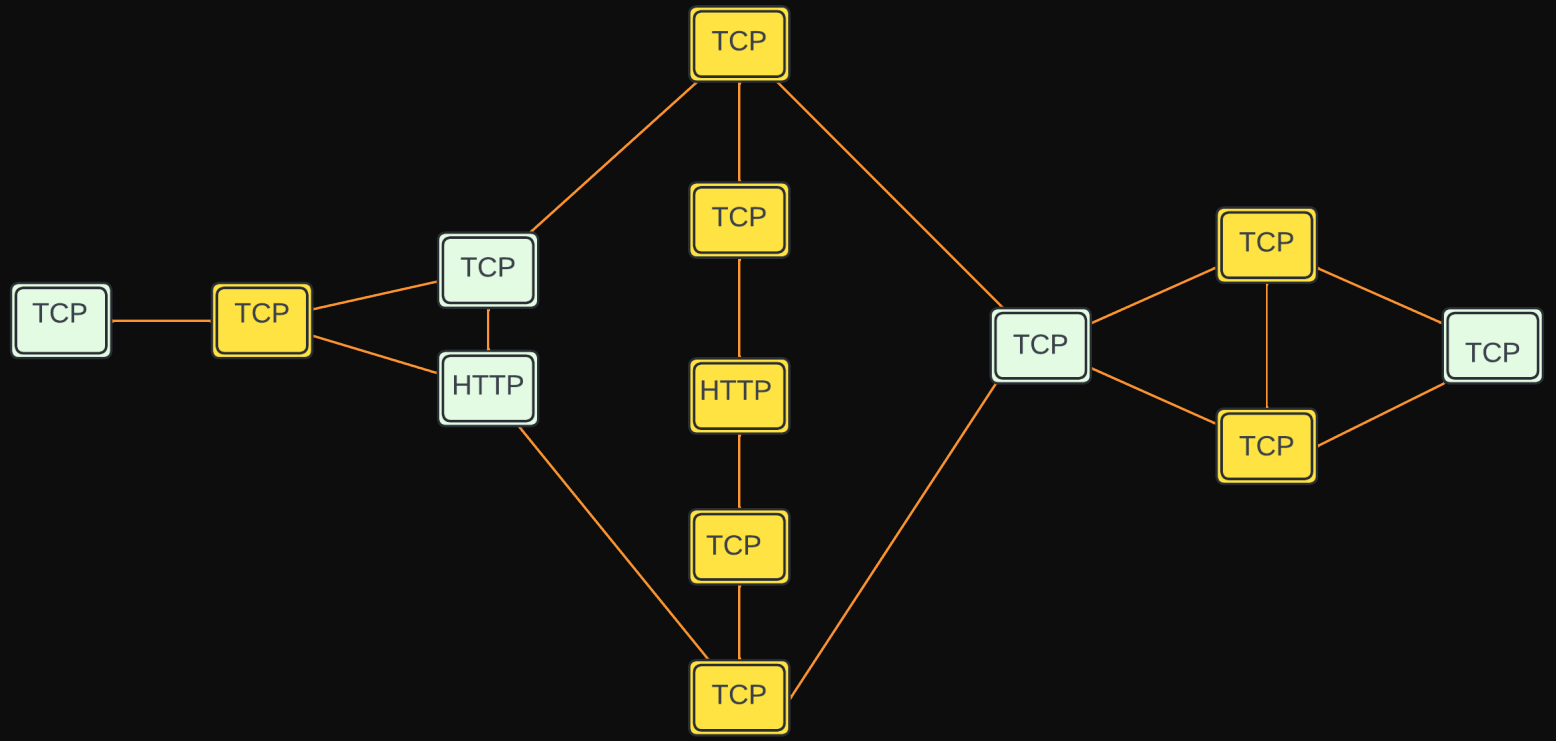
- ❑ Computational complexity
- ❑ Interpretability challenges

Using Packets as nodes

How is the network modeled?

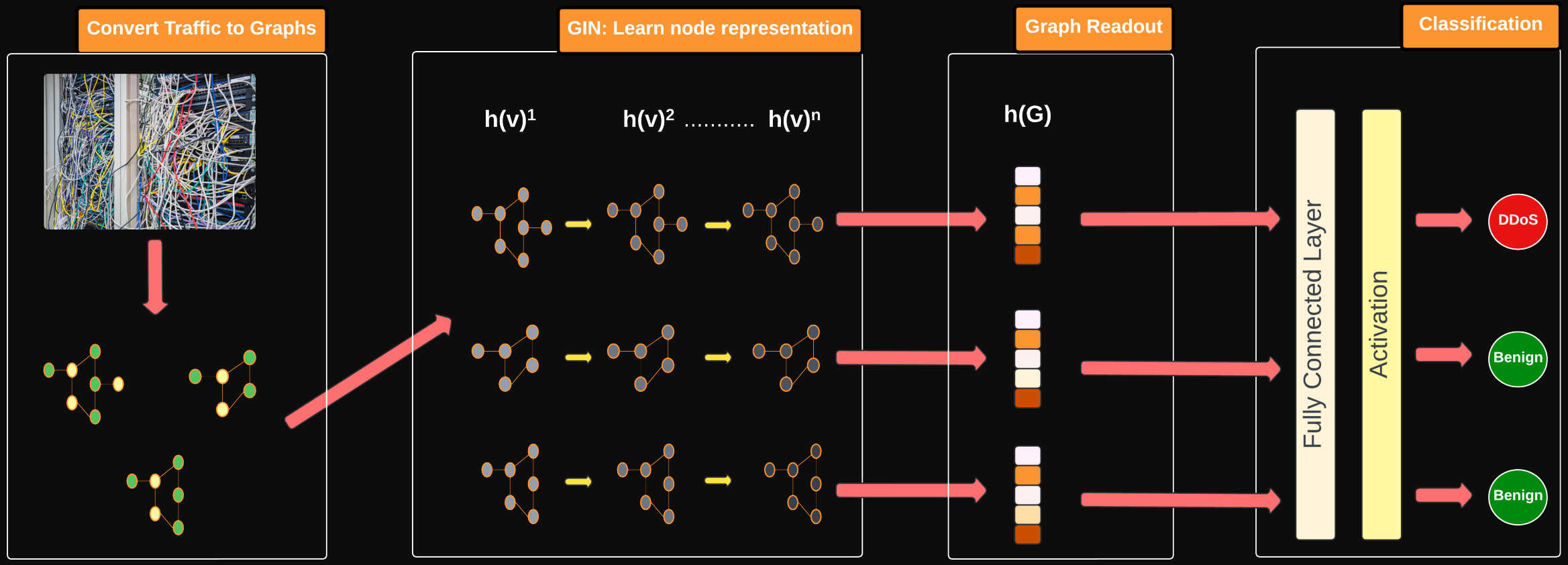
- Packets are **grouped** by source and destination IP.
- Packets are **sorted** by timestamp in ascending order.
- Node Creation: Packets become nodes.
  - Limited by pre-defined max number
  - Features: protocol type (e.g., TCP, UDP)
- Edge Types:
  - Between consecutive packets (same direction)
  - Between last packet of one direction and first of opposite

Using Packets as nodes



The endpoint traffic graph

Using Packets as nodes



Using Packets as nodes

RESULTS

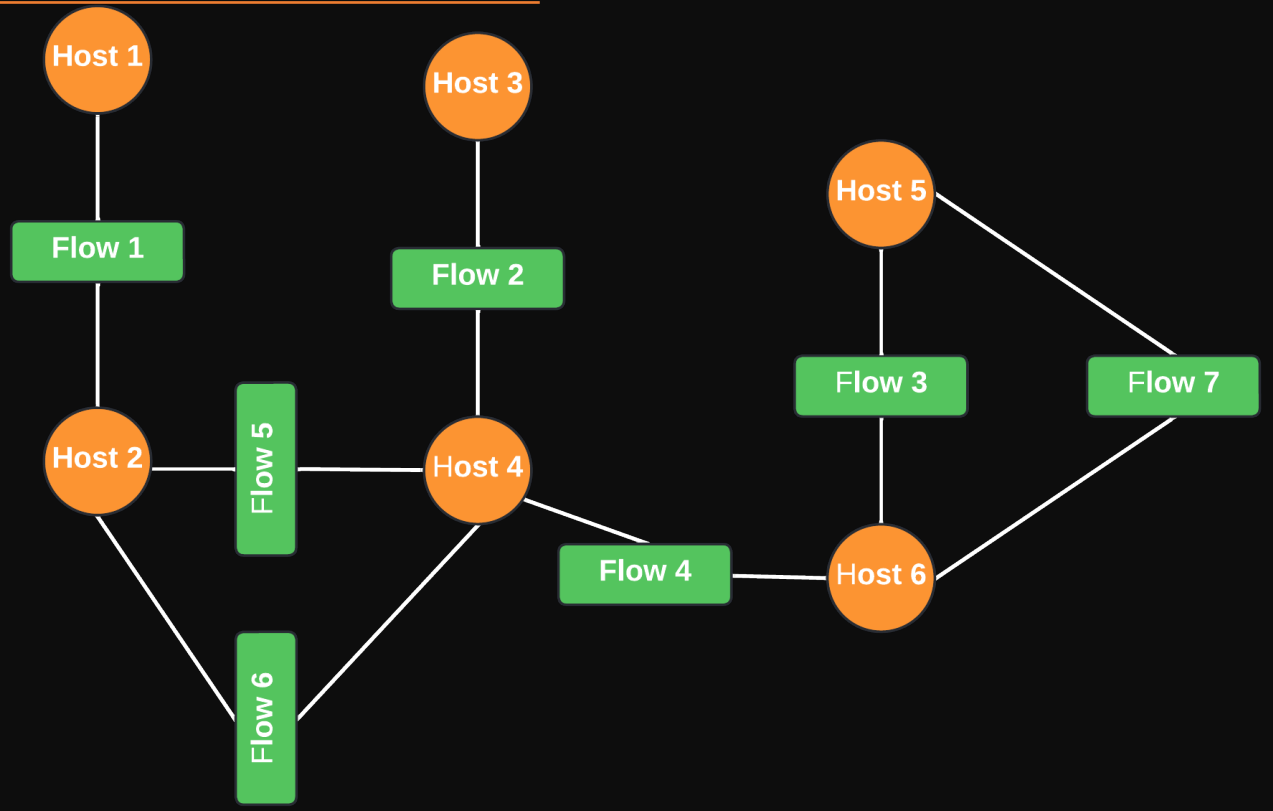
Datasets	Accuracy	Precision	Recall	F1
CIC-IDS2017	0.9959	0.9965	0.9953	0.9959
CIC-DOS2017	0.9751	0.9505	0.9407	0.9456

Using Traffic Flows as nodes

How is the network modeled?

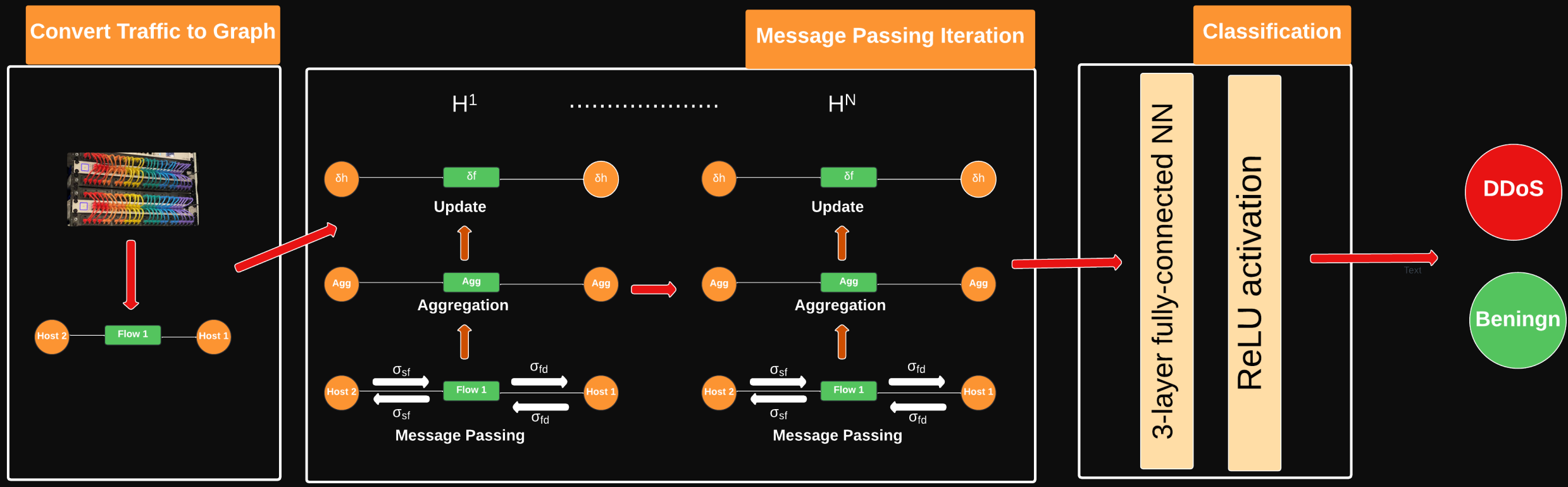
- ❑ Node Creation:
  - ❑ Host nodes: Represent source and destination Ips
  - ❑ Flow nodes: Represent individual network flows
  - ❑ Features:
    - ❑ Flow nodes: 80 features from the dataset (e.g., packet size, duration)
    - ❑ Host nodes: Initialized with all ones
- ❑ Edge Types:
  - ❑ Source-to-flow edges: Connect source host to flow
  - ❑ Flow-to-destination edges: Connect flow to destination host

Using Traffic Flows as nodes



Host-Connection Graph

Using Traffic Flows as nodes



Using Traffic Flows as nodes

RESULTS

Datasets	DoS GoldenEye	DosHulk	DoS slowloris	DoS Slowhttptest	DDoS
CIC-IDS2017	0.9959	0.9965	0.9953	0.9959	0.99

Accuracy over different attack classes

References

[1] Li, Yuzhen, et al. "Graphddos: Effective ddos attack detection using graph neural networks." 2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD). IEEE, 2022.

[2] Pujol-Perich, David, et al. "Unveiling the potential of graph neural networks for robust intrusion detection." ACM SIGMETRICS Performance Evaluation Review 49.4 (2022): 111-117.

Thank You