AI-Driven Intelligent Network Intrusion Detection System

Presented by jaemin Jeong

Mobile & Embedded System Lab.
Dept. of Computer Engineering
Kyung Hee Univ.

Contents

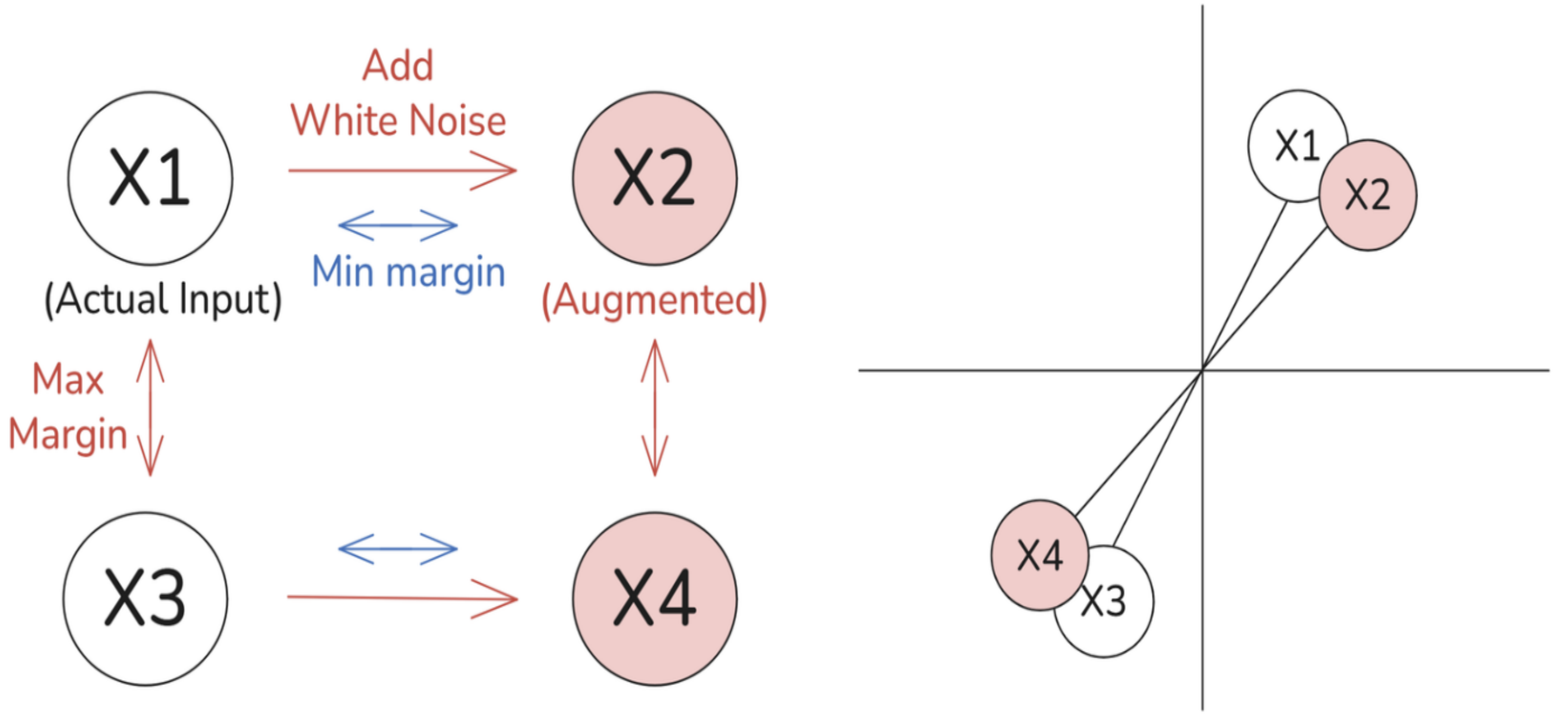
1. Multi-agent based Continual Representation learning

2. Toy example test results

3. Future Works

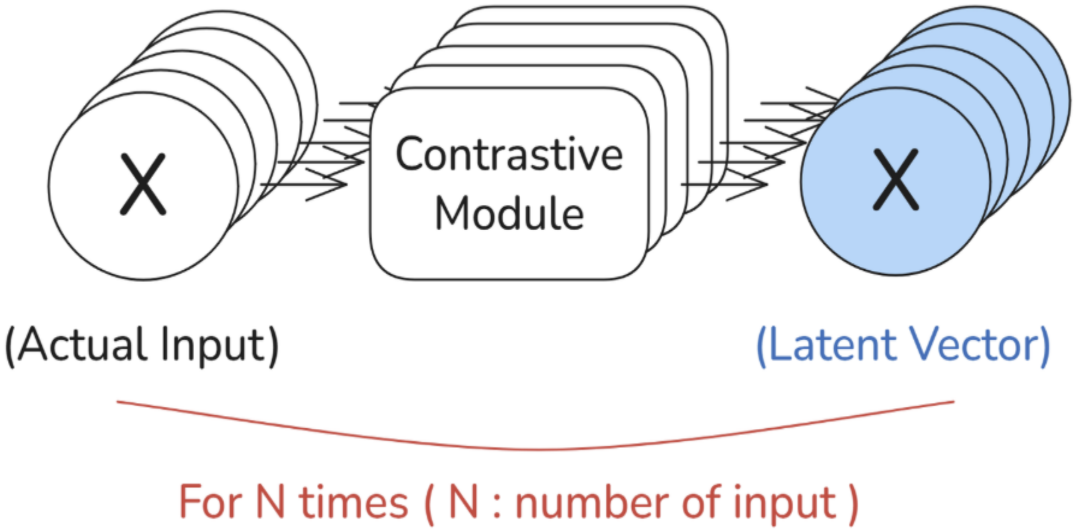
4. Discussion

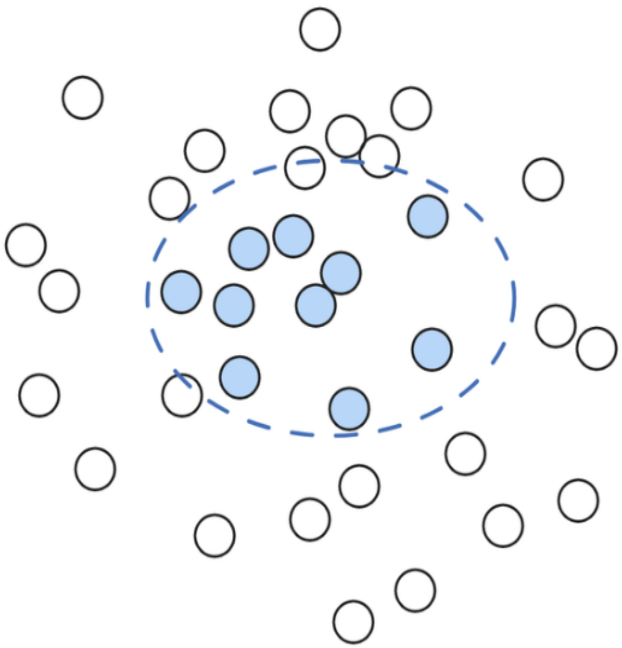
Unsupervised contrastive representation



- For unsupervised framework, we generate augmented input by adding **white noise** to actual input
- By contrastive learning, same samples (x1, x2) get closer, different sample (x1, x3) get further

Unsupervised Clustering Method





- Embed N input samples into latent vector using trained Contrastive Module
- To detect anomalies we utilize anomaly score based on Z-statistics $\mu \pm \text{zscore} \times \sigma$

1. Client side : Continual Representation Learning

Continual Learning : Adaptive EWC + Rehearsal buffer

$$\mathcal{L} = \mathcal{L}_{\text{const}} + \mathcal{L}^*_{\text{EWC}}$$

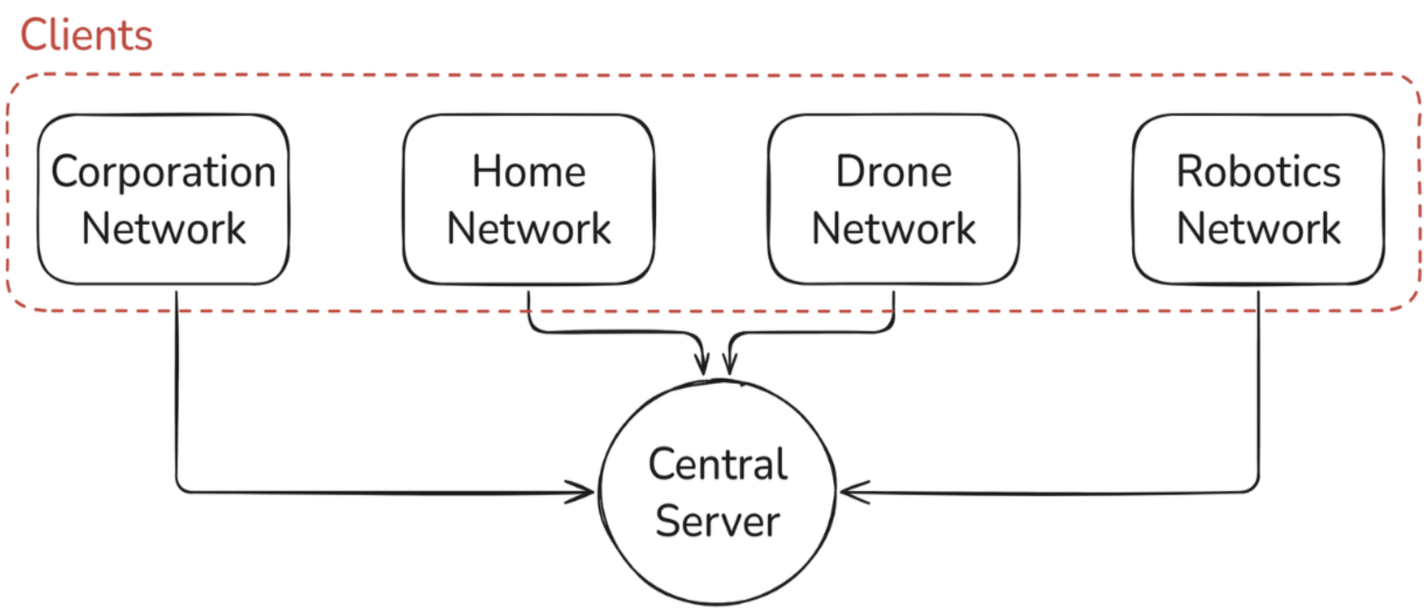
$$\mathcal{L}_{\text{EWC}} = \Sigma_i \frac{\lambda}{2} F_i(\theta_i - \theta^*_{A,i})^2$$

$$\mathcal{L}^*_{\text{EWC}} = \frac{\mathcal{L}_t}{\mathcal{L}_{t-1}} \Sigma_i \frac{\lambda}{2} F_i(\theta_i - \theta^*_{A,i})^2$$

- We used modified elastic weight consolidation to control the catastropic forgetting using loss ratio
- Also, Rehearsal buffer remains previous data samples into current data samples

1. Server side : Multi-Agent Based System

Federated learning framework : 1 Server, 4 Clients



- To implement the federated framework, we seperated 4 heterogenuous network into 4 clients
- For now central server just using default averaging method (to be changed into weighted sum)

2. Toy example test results

CIC-IDS17 Dataset (420,000 samples, 3 local epochs)

(Accuracy, FAR)	Phase (Benign Data Ratio, Attack Data Ratio)			
	1(0.9, 0.1)	2 (0.2, 0.8)	3 (0.1, 0.9)	4 (1.0, 0.0)
+ Adaptive EWC	61.3, 3.06	72.9, 2.39	63.5, 2.44	97.4, 2.51
+ EWC	65.6, 1.56	72.8, 2.61	47.9, 2.23	97.8, 2.42
+ Rehearsal buf	69.3, 2.55	72.4, 2.73	47.5, 1.76	97.7, 2.28
Transformer	66.4, 2.92	54.7, 3.44	49.3, 2.70	97.5, 2.50
MLP	48.1, 1.73	17.1, 3.50	24.7, 1.46	99.4, 0.5

3. Future Works

Develop toy example into real federated framework

- As-Is : Split single dataset into 4 pieces and seperate them into 4 federated clients
- To-Be : Seperate 4 different domain datasets into 4 federated clients (4 simulators)

Try another Contrastive Module

- As-Is : BYOL contrastive module (used in image classification domain)
- To-Be : Implement additional contrastive module (InfoMCE) and compare both

Advanced Federated-learning aggregation method

- As-Is : Defaul averaging method
- To-Be : Weighted sum (Calculate weights using convex optimization or simple loss ratio diff)

4. Discussion

Adding Generative Module?

- We're trying unsupervised module which we don't know which is attack to be augmented
- I'm not sure about - which part should be the most appropriate part (doing more paper review)

About Evaluation Methods

1. Simple Evaluation : Default metric based evaluation (train - test dataset split)
2. Federated Evaluation : Learn from client-side dataset → Test from server-side inference
3. Continual Evaluation : Learn attack A → attack B (keep measuring metrics)

- We built 4 different simulators which can do actual attack simulation on these things.
- We're trying to build automated attack scripts including various types of attack whether it's in training dataset or not.
- For this situation, which method would be appropriate and do we need simulator in evaluation?

Q & A

http://mesl.khu.ac.kr