Generative Adversarial Networks Adaptation Using Transfer Learning

Fernando Nakayama

Research Advisor: Dr. Michele Nogueira Lima

Center for Computational Security sCience (CCSC)

Belo Horizonte – Brazil

March 10, 2023

www.ccsc-research.org

DA-GAN: Domain Adaptation for Generative Adversarial Networks-assisted Cyber Threat Detection

Hien Do Hoang*†, Do Thi Thu Hien*†, Thai Bui Xuan*†, Tri Nguyen Ngoc Minh*†, Phan The Duy*†, and Van-Hau Pham*†

2022 RIVF International Conference on Computing and Communication Technologies (RIVF)

Network Traffic Prediction Based on LSTM and Transfer Learning

XIANBIN WAN[1], HUI LIU[1], HAO XU[1], AND XINCHANG ZHANG[2], (Senior Member, IEEE)
[1]Shandong Computer Science Center (National Supercomputer Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences), Jinan 250014, China
[2]College of Computer Science and Technology, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250306, China
Corresponding author: Xianbin Wan (150691180 31@163.com)

IEEE Access, 2022 - ieeexplore.ieee.org

DA-GAN: Domain Adaptation for Generative Adversarial Networks-assisted Cyber Threat Detection

Hien Do Hoang*†, Do Thi Thu Hien*†, Thai Bui Xuan*†, Tri Nguyen Ngoc Minh*†,
Phan The Duy*†, and Van-Hau Pham*†

2022 RIVF International Conference on Computing and Communication Technologies (RIVF)

DA-GAN: Domain Adaptation for Generative Adversarial Networks-assisted Cyber Threat Detection

Context

- ML-assisted IDS

Problem

- Labeled data is scarce
- Private data
- Unbalanced data

Contribution

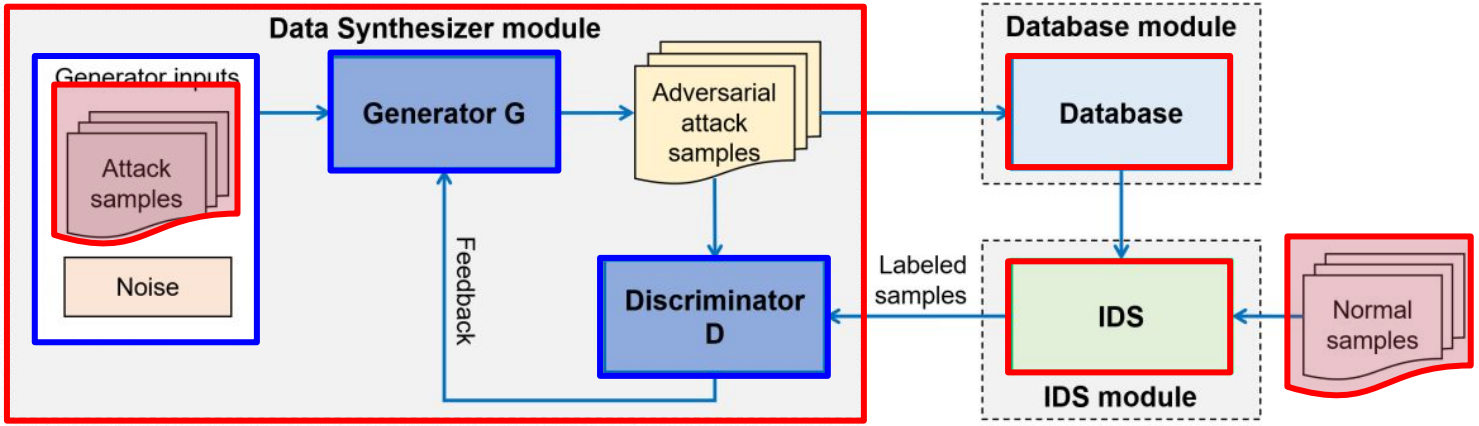
- DA-GAN (Domain Adaptation + Generative Adversarial Networks)

Ben-David, Shai, et al. "A theory of learning from different domains." *Machine learning* 79 (2010): 151-175.

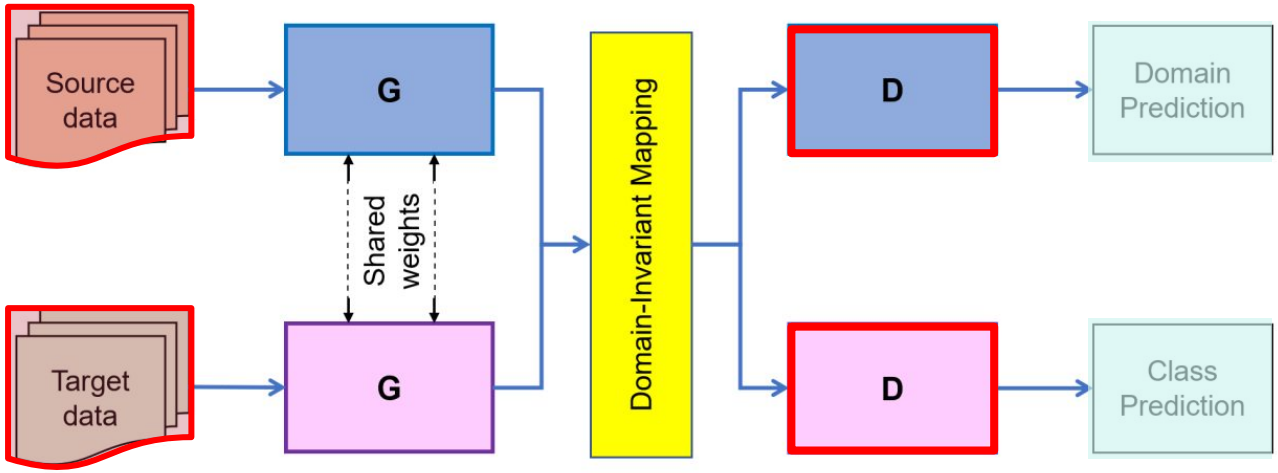
Dataset preparing and preprocessing

- Dataset Preparing and Preprocessing
  - Dataset
  - Preprocessing
    - Feature selection
    - Data numeration
    - Value normalization

DA-GAN design



Domain adaptation for DA-GAN



IMPLEMENTATION AND EXPERIMENTS

- **Datasets**
  - **CIC-IDS-2018**
  - **68 numeric features after feature selection and normalization**
  - **40%, 40%, and 20%**

Class Label	Number of records
Benign	2,856,035
Bot	286,191
Brute Force	513
DoS	1,289,544
Infiltration	93,063
SQL Injection	53

IMPLEMENTATION AND EXPERIMENTS

- Google Colab
- 3 GAN variants - WGAN, WGAN-GP, WGAN-GP-TTUR
- IDS - Linear Regression (LR), Support Vector Machine (SVM), Random Forest (RF), K-nearest Neighbors (KNN) (scikit-learn)

Operation	Value
Optimizer	RMSProp
Batch size	64
Weight clipping	0.01
Number of epoch	50
Learning rate in WGAN and WGAN-GP	
Generator G	0.0001
Discriminator D	0.0001
Learning rate in WGAN-GP-TTUR	
Generator G	0.0001
Discriminator D	0.0002
Num. Iteration of D	5

IMPLEMENTATION AND EXPERIMENTS

- **Applying Domain Adaptation (DA)**
  - **WGAN (same hyperparameters)**

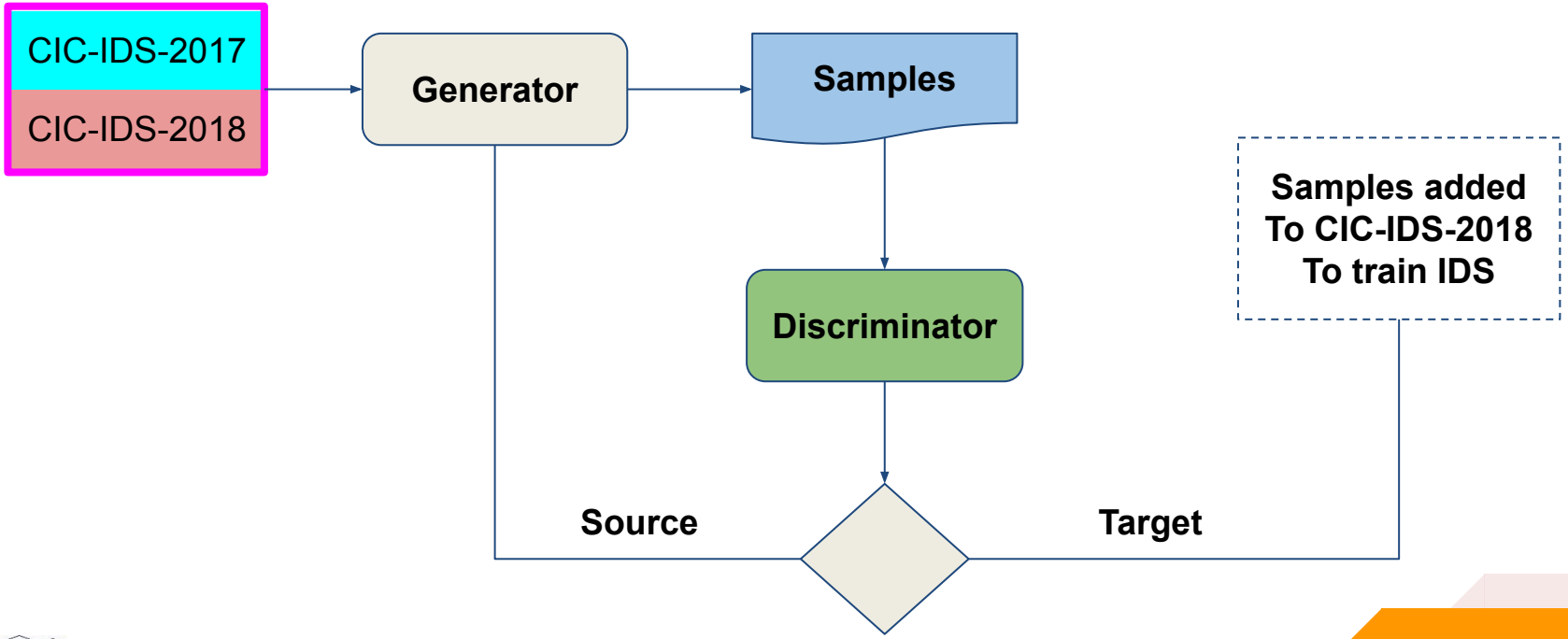
CIC-IDS-2017 (source domain)

Class Label	Number of records
Benign	2,359,289
FTP-Patator	7,938
SSH-Patator	5,897
DoS-GoldenEye	10,293
DoS-Hulk	231,073
DoS-Slowhttp	5,499
DoS-Slowloris	5,796
Heartbleed	11
Web-Attack-BruteForce	1,507
Web-Attack-SQLInjection	21
Web-Attack-XSS	652
Infiltration	36
Bot	1,966
PortScan	158,930
DDoS	41,835

CIC-IDS-2018 (target domain)

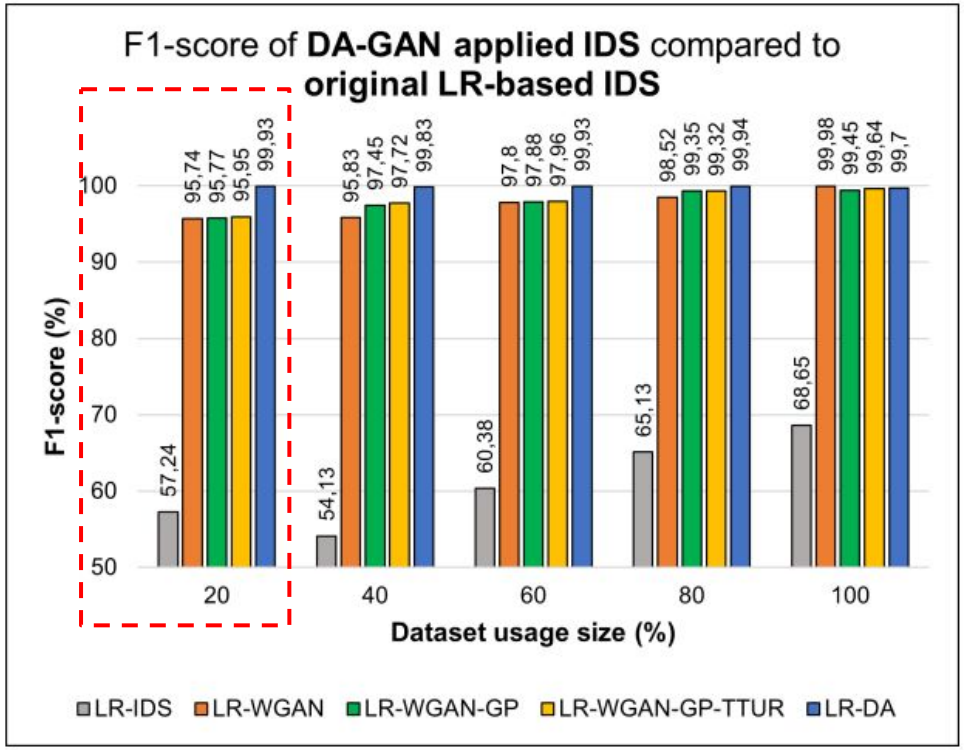
Class Label	Number of records
Benign	2,856,035
Bot	286,191
Brute Force	513
DoS	1,289,544
Infiltration	93,063
SQL Injection	53

DA-GAN design



Training dataset size (%)	Model	LR			SVM			RF			KNN		
		F1	AUC	Recall	F1	AUC	Recall	F1	AUC	Recall	F1	AUC	Recall
20	Original IDS	57.24	58.36	54.86	59.25	59.92	58.25	60.24	60.36	60.86	56.53	58.54	56.86
	WGAN	95.74	99.93	95.13	97.81	99.93	97.29	98.48	99.93	97.65	98.27	99.93	98.82
	WGAN-GP	95.77	98.33	93.26	96.01	98.52	96.89	98.56	98.33	98.85	97.77	98.33	97.26
	WGAN-GP-TTUR	95.95	99.93	93.54	96.12	99.93	96.26	99.09	99.93	99.54	97.85	99.92	97.54
	Domain Adaptation	99.93	99.93	99.93	99.93	99.93	99.93	99.93	99.93	99.93	99.93	99.93	99.91
40	Original IDS	56.13	57.65	56.32	58.24	57.65	58.32	62.13	63.65	60.86	58.13	57.65	58.32
	WGAN	95.83	99.93	96.33	95.83	99.93	97.37	98.83	99.93	97.65	96.83	98.93	96.33
	WGAN-GP	97.45	99.93	96.63	98.05	99.93	98.28	98.26	99.93	98.85	98.47	99.93	98.66
	WGAN-GP-TTUR	97.72	99.93	97.33	98.41	99.93	96.9	98.23	99.93	99.54	97.72	99.93	97.33
	Domain Adaptation	99.83	99.93	99.92	99.83	99.93	99.92	99.98	99.93	99.93	99.63	99.93	99.92
60	Original IDS	60.38	60.54	59.56	60.98	60.54	59.65	63.26	62.54	60.52	65.13	65.36	64.53
	WGAN	97.8	98.26	98.24	98.31	98.56	99.33	98.21	98.26	98.33	98.6	98.26	99.24
	WGAN-GP	97.88	99.93	98.36	97.53	99.93	98.26	98.88	99.96	98.63	98.86	99.93	98.37
	WGAN-GP-TTUR	97.96	99.93	97.56	98.49	99.93	98.86	97.96	99.96	98.33	98.93	99.93	98.63
	Domain Adaptation	99.93	99.93	99.93	99.95	99.93	99.98	99.97	99.93	99.92	99.98	99.93	99.96
80	Original IDS	65.13	64.53	64.89	64.13	64.53	63.89	65.14	64.57	65.89	65.17	64.53	65.79
	WGAN	98.52	100	98.78	98.58	99.86	97.36	98.49	99.93	98.86	98.69	99.53	98.26
	WGAN-GP	99.35	99.93	99.56	98.62	99.93	98.13	99.15	99.93	99.93	99.65	99.93	99.56
	WGAN-GP-TTUR	99.32	99.93	99.36	99.15	99.93	99.63	99.32	99.93	99.36	99.38	99.93	99.38
	Domain Adaptation	99.94	99.93	99.93	99.96	99.93	99.96	99.94	99.93	99.93	99.98	99.93	99.97
100	Original IDS	68.65	68.32	69.36	70.65	70.33	69.36	70.65	70.33	69.83	73.62	72.33	73.38
	WGAN	99.98	100	99.96	99.68	100	99.36	99.64	99.93	99.29	99.98	100	99.96
	WGAN-GP	99.45	99.93	99.26	99.64	99.93	99.29	99.32	99.93	99.36	99.45	99.93	99.26
	WGAN-GP-TTUR	99.64	99.93	99.95	99.76	99.93	99.59	99.63	99.93	99.96	99.64	99.93	99.95
	Domain Adaptation	99.7	100	99.85	99.77	100	99.58	99.9	100	99.65	99.76	100	99.88

Results


F1-score of DA-GAN applied IDS compared to original LR-based IDS
F1-score (%)
100
90
80
70
60
50
57,24
95,74
95,77
95,95
99,93
54,13
95,83
97,45
97,72
99,83
60,38
97,8
97,88
97,96
99,93
65,13
98,52
99,35
99,32
99,94
68,65
99,98
99,45
99,64
99,7
20
40
60
80
100
Dataset usage size (%)
LR-IDS
LR-WGAN
LR-WGAN-GP
LR-WGAN-GP-TTUR
LR-DA

Discussion - Paper Conclusion

- GAN is trained in the source dataset
- New samples are generated for target dataset
- Effective GAN + DA architecture
- Improves performance of ML-Based IDS

Discussion - Considerations

The preprocessing process mentioned in **Section III-A2** to have data in the proper structure to feed to ML-DL models. Whereby, each preprocessed record consists of 68 numeric features after feature selection and normalization. Moreover, based on the work of Simon et al. [23] and our main focus on DoS attack, our proposed system attempts to stay away from functional features which are Flow Duration, Active Mean, Average Packet Size, Packet Length Std, Flow IAT Mean, PSH Flag Count and Idle Max in adversarial sample creation to retain the function of attack records. After processing the

fernandonakayama@ufpr.br