BGP Features and Classification of Internet Worms and Ransomware Attacks

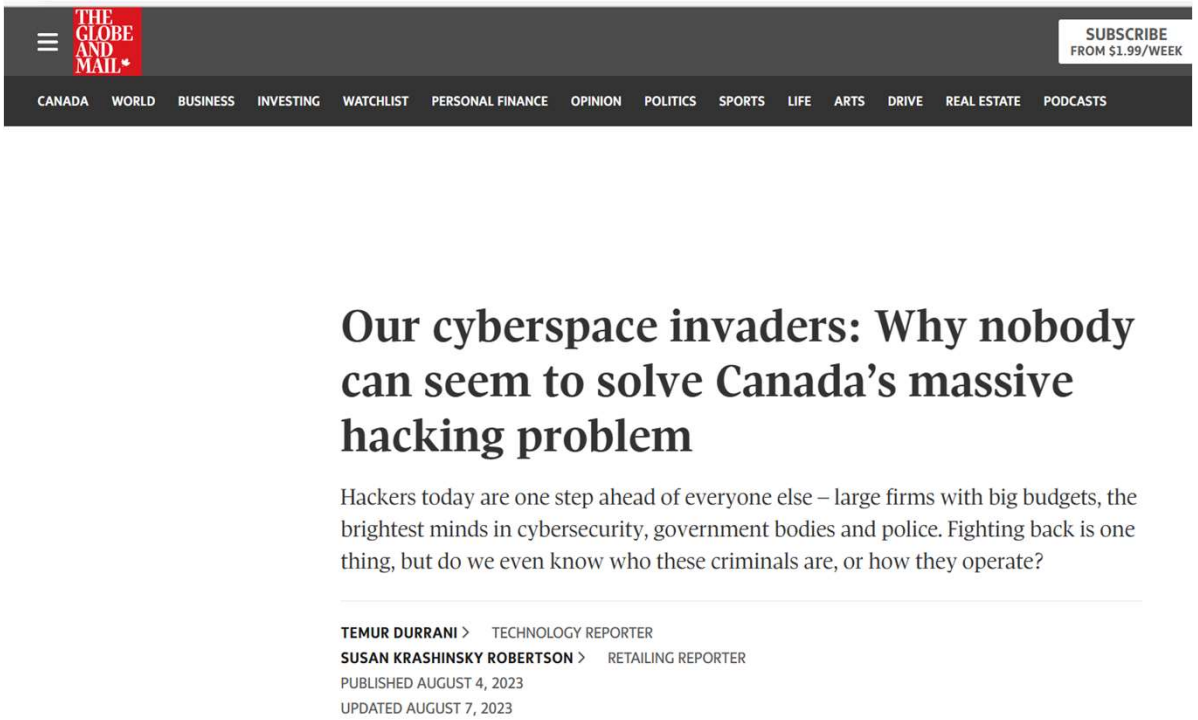
Hardeep Kaur Takhar and Ljiljana Trajković
Communication Networks Laboratory
http://www.sfu.ca/~ljilja/cnl
Simon Fraser University
Vancouver British Columbia, Canada

IEEE SMC 2023
Oahu, Hawaii

Roadmap

- Introduction
- Datasets
- Feature analysis and machine learning
  - Feature selection
  - K-S test for estimating BGP feature distributions
  - Classification using GBDT algorithms
- Experiments and performance results
- Conclusions and references

Cyberattacks



https://www.theglobeandmail.com/business/article-cybersecurity-cybercrime-hack-canada/

Introduction

- New network applications and generated traffic:
  - result in increased system vulnerabilities
  - exposed networks to security threats
- Traffic anomalies result from malicious attacks leading to unusual traffic patterns:
  - significant disruptions in communication networks
  - performance-related:
    - file server failures, network congestion, packet flooding
  - security-related:
    - viruses, worms, denial of service (DoS) and distributed DoS attacks, trojans, rootkits, and ransomware attacks

Introduction

- Worms:
  - compromise systems by excessively consuming network resources and make them inaccessible to legitimate users
  - Code Red (2001), Nimda (2001), Slammer (2003)
- Power outages:
  - Moscow (2005) and Pakistan (2021) blackouts
- Denial of Service (DoS) and DDoS attacks:
  - CIC-IDS 2017, CSE-CIC-IDS 2018, CIC-DDoS 2019
- Ransomware:
  - use advanced cryptography techniques to encrypt data and demand ransom
  - WannaCrypt (2017), WestRock (2021)

Border Gateway Protocol (BGP)

- De facto interdomain Internet routing protocol
- BGP messages:
  - Open
  - Keepalive
  - Update:
    - protocol status and configurations
    - critical information about the network connectivity
  - Notification

RFC 1771 - A border gateway protocol 4 (BGP-4).
[Online]. Available: https://datatracker.ietf.org/doc/html/rfc1771. Accessed: Aug. 2023.

Roadmap

- Introduction
- Datasets
- Feature analysis and machine learning
  - Feature selection
  - K-S test for estimating BGP feature distributions
  - Classification using GBDT algorithms
- Experiments and performance results
- Conclusions and references

Datasets

- Generated using BGP update messages
- Collection sites:
  - Réseaux IP Européens (RIPE):
    - Routing Information Service project by RIPE Network Coordination Centre
  - Route Views:
    - University of Oregon project

RIPE Network Coordination Centre: About us. [Online]. Available: https://www.ripe.net/about-us/. Accessed: Apr. 2023.

RIPE NCC. [Online]. Available: https://www.ripe.net . Accessed: Apr. 2023.

University of Oregon Route Views project. [Online]. Available: http://www.routeviews.org . Accessed: Apr. 2023.

Datasets

- Regular: two days prior and two days after the attack
- Anomalous: reported days of the attack
- Each row represents one minute of the collected data
- 37 extracted features:
  - volume and AS-path
- Binary classification:
  - regular: 0
  - anomaly: 1
- Training and test datasets contain: 60 % and 40 % of the anomalies

Border Gateway Protocol Routing Records from Réseaux IP Européens (RIPE) and BCNET. [Online]. Available: http://ieee-dataport.org/1977. Accessed: Aug. 2023.

RIPE: BGP Datasets

Dataset	Regular (min)	Anomaly (min)	Regular (training)	Anomaly (training)	Regular (test)	Anomaly (test)	Start 00:00:00	End 23:59:59
Code Red	6,600	600	3,679	361	2,921	239	17.07.2001	21.07.2001
Nimda	7,308	1,301	3,673	827	3,635	474	16.09.2001	21.09.2001
Slammer	6,331	869	3,210	530	3,121	339	23.01.2003	27.01.2003
WannaCrypt	5,760	5,760	2,880	3,420	2,880	2,340	10.05.2017	17.05.2017
WestRock	5,832	10,008	2,952	6,008	2,880	4,000	21.01.2021	31.01.2021

Z. Li, A. L. Gonzalez Rios, and Lj. Trajkovic, "Machine learning for detecting the WestRock ransomware attack using BGP routing records," *IEEE Commun. Mag.*, vol. 61, no. 3, pp. 20–26, Mar. 2023

Z. Li, A. L. Gonzalez Rios, and Lj. Trajkovic, "Machine learning for detecting anomalies and intrusions in communication networks," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 7, pp. 2254–2264, July 2021.

BGP Update Messages: Features

Feature Number	Name	Category
1	Number of announcements	volume
2	Number of withdrawals	volume
3/4	Number of announced/withdrawn NLRI prefixes	volume
5/6/7	Average/maximum/average unique AS-path length	AS-path
8/10	Number of duplicate announcements	volume
9	Number of implicit withdrawals	volume
11/13	Maximum/average edit distance	AS-path
12	Arrival rate	volume
14-23/24-33	Maximum AS-path length/edit distance	AS-path
34/35/36	Number of IGP/EGP/incomplete packets	volume
37	Packet size	volume

NLRI: Network Layer Reachability Information

Roadmap

- Introduction
- Datasets
- Feature analysis and machine learning
  - Feature selection
  - K-S test for estimating BGP feature distributions
  - Classification using GBDT algorithms
- Experiments and performance results
- Conclusions and references

Feature Selection

- Statistical approaches:
  - Correlation coefficients:
    - Pearson (ρ): linear relationships
    - Spearman (r_s): non-linear relationships
  - $\rho, r_s \in [-1, 1]$:
    - +1: strong-positive; -1: strong-negative; 0: no relationship
- Supervised machine learning:
  - Random forests
  - Extra-trees

K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: The MIT Press, 2012.

Feature Selection: Random Forests

- Employ bootstrap aggregation (bagging) to generate multiple uncorrelated decision trees
- Bagging: uses bootstrapping resampling technique to uniformly sample data using replacement
  - data point might appear multiple times in a given training dataset
- Decision trees in random forests are generated using a random approach
  - to select a subset of features and threshold values for splitting
- Quality of a split is measured using Gini impurity
- Each model is independently trained in parallel using samples selected by bagging
- After decision trees are built, each model makes a prediction
- Outcome with a majority vote is selected as the output

L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Jan. 2001.

Feature Selection: Extra-Trees

- Extremely Randomized Trees (extra-trees):
  - derived from random forest
  - faster execution time
  - each decision tree is trained using a complete dataset without resampling
  - split point for each decision tree is selected randomly
  - feature scores are calculated based on Gini importance
- Experiments indicate that extra-trees are a better approach for selecting important features

P. Geurts, D. Ernst, and L. Wehenkel, “Extremely randomized trees,” *Mach. Learn.*, vol. 63, no. 1, pp. 3–42, Apr. 2006.

Feature Analysis: Goodness of Fit Test

- Goodness of fit Kolmogorov–Smirnov (K–S) test:
  - compares sampled data distribution with the reference PDFs
- Probability distributions selected to estimate BGP features:
  - Gaussian (normal), exponential, gamma
  - Heavy-tailed:
    - Weibull, Rayleigh, Burr, t Location-Scale, log-normal, log-logistic
- Traffic traces in communication networks often follow heavy-tailed distributions

Y. Dodge, *The Concise Encyclopedia of Statistics*. New York, NY: Springer New York, 2008, pp. 283–287.

N. T. Thomopoulos, *Statistical Distributions Applications and Parameter Estimates*. Cham, Switzerland: Springer Nature, 2017.

A. Alzaatreh, C. Lee, and F. Famoye, "A new method for generating families of continuous distributions," *METRON*, vol. 71, pp. 63–79, June 2013.

Classification Algorithms: Ensemble Learning

- Sequentially combines models to generate an optimal model
- Gradient boosting decision tree (GBDT) algorithms:
  - variants of gradient boosting machines (GBM)
  - employ functional gradient descent to optimize the loss function
- GBDT models: trained by sequentially adding base learners (decision trees) to achieve the minimum loss

J. Friedman, "Greedy function approximation: a gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Apr. 2001.

GBDT Classification Algorithms

- eXtreme gradient boosting (XGBoost):
  - asymmetrically level-wise
- Light gradient boosting (LightGBM):
  - asymmetrically leaf-wise
- Categorical boosting (CatBoost):
  - symmetric

T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, San Francisco, CA, USA, Aug. 2016, pp. 785–794.

G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.- Y. Liu, "LightGBM: a highly efficient gradient boosting decision tree," in *Proc. Int. Conf. Neural Inform. Process. Syst.*, Long Beach, CA, USA, Dec. 2017, pp. 3146–3154.

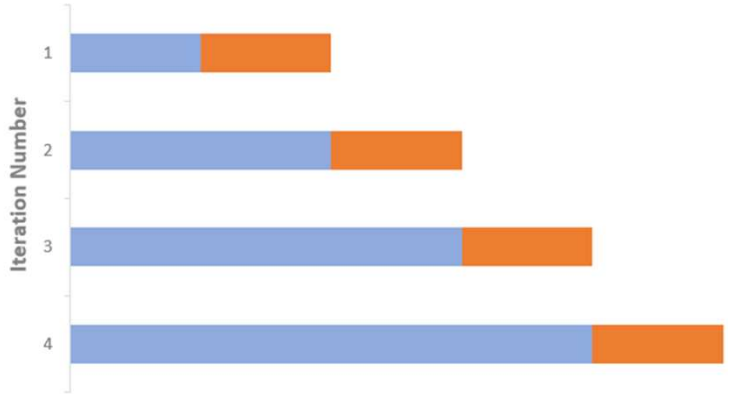
L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical features," in *Proc. Int. Conf. Neural Inform. Process. Syst.*, Montreal, QC, Canada, Dec. 2018, pp. 6639–6649.

Roadmap

- Introduction
- Datasets
- Feature analysis and machine learning
  - Feature selection
  - K-S test for estimating BGP feature distributions
  - Classification using GBDT algorithms
- Experiments and performance results
- Conclusions and references

Cross-Validation

- Time series split:
  - variation of 10-fold cross validation
  - training (blue) and test (orange) datasets
  - successive training datasets are concatenated over time
  - Maintains a time sequence of sequential data



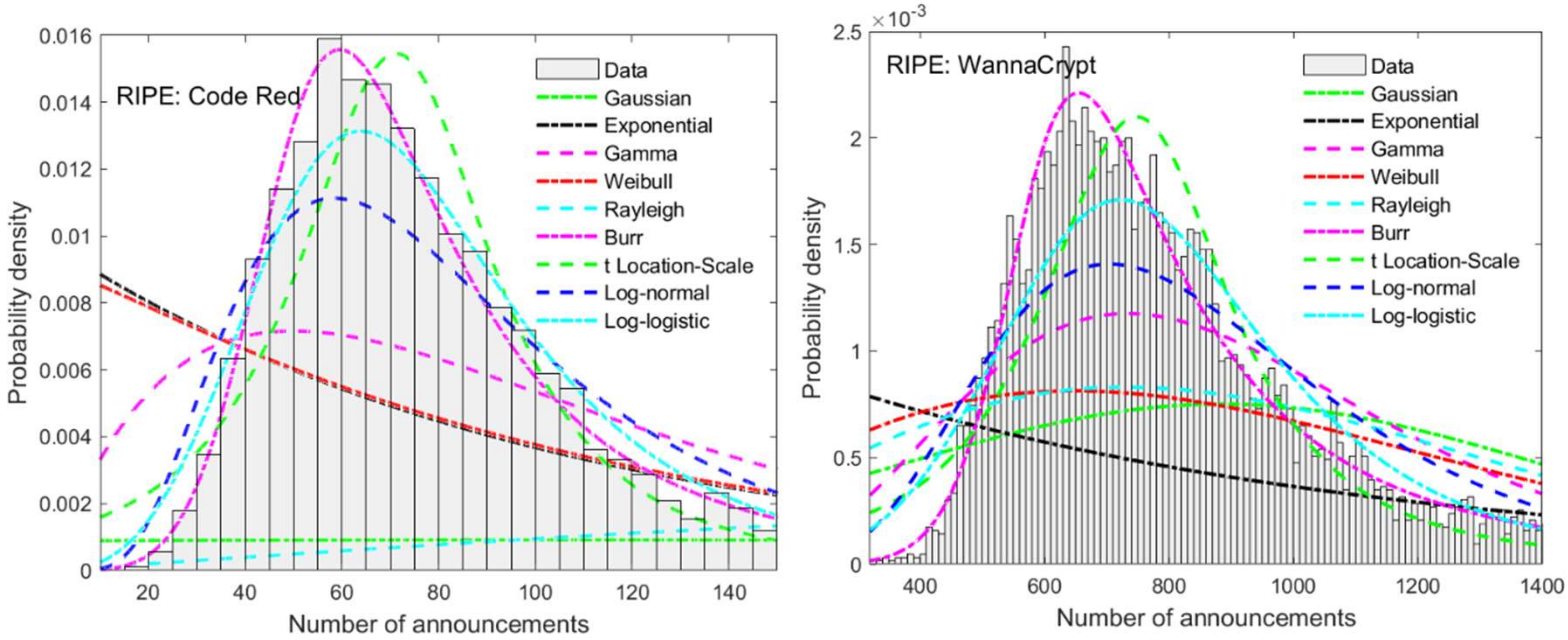
Feature Selection: Extra-Trees

- 10-fold time-series split cross-validation experiments performed based on accuracy and F-Score
- Model hyperparameters:
  - number of estimators = 500; maximum tree depth = 20

Dataset	Feature numbers in order of importance
Code Red	34, 1, 4, 3, 12, 2, 9, 37, 36, 8, 10, 13, 5, 7, 35, 6
Nimda	1, 34, 3, 4, 9, 36, 12, 37, 8, 23, 10, 2, 13, 7, 11, 5
Slammer	36, 1, 9, 34, 10, 8, 3, 4, 2, 20, 11, 12, 6, 13, 5, 7
WannaCrypt	4, 8, 2, 3, 10, 37, 1, 34, 36, 9, 12, 13, 35, 11, 6, 7
WestRock	8, 9, 3, 37, 2, 1, 36, 34, 10, 4, 12, 35, 13, 6, 11, 7

K-S Goodness of Fit Test: PDFs

- Nine probability distributions:
  - top 10 important BGP features selected using extra-trees



K-S Goodness of Fit Test: PDF Candidates

- Selected based on visual inspection:
  - Burr, t Location-Scale, log-normal, log-logistic
- Statistical measures: h, p-value, k, c

Distribution Feature Number	α = 0.10	α = 0.05	α = 0.01
Burr		p-value = 0.292473, k = 0.015371	
F3: Code Red			
h	0	0	0
c	0.019214	0.021325	0.025565
Log-normal		p-value = 0.292473 k = 0.015371	
F9: Nimda			
h	0	0	0
c	0.018207	0.020208	0.024225
Burr		p-value = 0.034104, k = 0.023285	
F3: Slammer			
h	1	1	0
c	0.019968	0.022162	0.026569
Burr		p-value = 0.376152, k = 0.011466	
F3: WannaCrypt			
h	0	0	0
c	0.015393	0.017084	0.020479
Log-logistic		p-value = 0.284391, k = 0.010407	
F4: WestRock			
h	0	0	0
c	0.012911	0.014329	0.017176

Feature Probability Distributions

- Accepted PDFs based on null hypothesis:
  - highlighted distributions: p-value ≥ α = 0.05

Dataset	Distribution	Features
Code Red	Burr	F34, F1, F3, F9, F37
Nimda	Burr/Log-normal/Log-logistic	F9
Slammer	Burr	F3
WannaCrypt	Burr	F4, F3, F10, F1, F34, F36, F9
WestRock	Burr	F9, F4

Fitting Distributions: Parameters

Dataset	Feature	Parameters		
Burr		α	c	k
Code Red	F34	48.7857	4.98972	0.47655
	F1	56.9317	5.31064	0.45235
Nimda	F9	92.1486	1.80949	0.98291
Slammer	F3	57.7592	3.15328	3.15328
WannaCrypt	F4	113.344	4.70707	4.70707
	F10	89.3155	2.92279	0.58448
WestRock	F4	613.496	6.07794	0.72901
Log-normal		μ	σ	
Nimda	F9	4.54569	0.97609	
Log-logistic		μ	σ	
Nimda	F9	4.53806	0.55604	
WestRock	F4	6.50227	0.18627	

K-S Goodness of Fit Test: Common Features

- Code Red and WannaCrypt datasets:
  - common features (F34, F1, F3) follow the Burr distribution indicate similarities between the two datasets
  - WannaCrypt being a cryptoworm propagates through a network using similar self-replication and self-propagation techniques employed by worms
- Code Red and WestRock datasets:
  - number of implicit withdrawals (F9) follows the Burr distribution
    - number of newly advertised AS-paths for the announced NLRI prefixes
  - indicates that during the attack traffic may have been re-routed through desired AS-paths by the attacker
- Code Red, Nimda, Slammer worm datasets:
  - no common features with accepted null hypotheses for a given PDF

NLRI: Network Layer Reachability Information

Machine Learning Models: Hyperparameters

- GBDT models: best performing hyperparameters using 37, 16, and 8 features:
  - based on accuracy and F-Score
  - time series split
    10-fold cross-validation

		Accuracy and F-Score	
Dataset	**Algorithm**	**Number of Estimators**	**Learning Rate**
	XGBoost	10	0.01
Code Red	LightGBM	10	0.01
	CatBoost	10	0.01
Nimda	XGBoost	260	0.01
	LightGBM	280	0.01
	CatBoost	240	0.10
Slammer	XGBoost	140	0.05
	LightGBM	170	0.01
	CatBoost	60	0.10
WannaCrypt	XGBoost	270	0.10
	LightGBM	130	0.10
	CatBoost	280	0.10
WestRock	XGBoost	330	0.10
	LightGBM	50	0.10
	CatBoost	170	0.05

GBDT Models: Best Performance

Dataset	Number of features	Training time (s)			Accuracy (%)			F-Score (%)		
		XGBoost	LightGBM	CatBoost	XGBoost	LightGBM	CatBoost	XGBoost	LightGBM	CatBoost
	37	0.0470	0.0425	0.2110	96.84	92.41	97.28	78.54	0.00	81.30
Code Red	16	0.0262	0.0253	0.0525	96.84	92.41	97.22	78.54	0.00	81.03
	8	0.0231	0.0272	0.0468	96.90	92.41	96.58	80.32	0.00	75.78
	37	1.0583	0.4607	2.4636	80.58	81.67	82.14	39.08	40.94	42.11
Nimda	16	0.6187	0.4359	2.3066	80.58	81.46	81.97	39.08	40.56	41.97
	8	0.4749	0.3122	0.7893	80.24	80.99	80.65	39.58	39.32	40.09
	37	0.4645	0.3848	0.2644	93.76	93.06	94.08	55.37	46.67	58.58
Slammer	16	0.2710	0.2114	0.1824	93.55	92.95	93.15	53.05	45.05	47.91
	8	0.1968	0.1599	0.1652	93.41	92.75	93.09	51.07	42.30	47.01
	37	1.9838	0.3276	1.7270	60.48	59.52	60.02	61.43	60.81	61.30
WannaCrypt	16	2.3483	0.3074	1.3041	60.13	59.66	59.90	61.05	60.81	61.43
	8	0.7880	0.1973	1.0541	61.03	60.54	60.13	61.95	61.47	61.75
	37	3.3058	0.1643	2.8835	57.79	57.35	56.31	71.48	71.07	70.32
WestRock	16	1.7448	0.1480	1.1840	57.73	57.53	56.58	71.33	71.26	70.90
	8	0.9798	0.0907	1.4420	59.56	58.02	56.38	72.96	71.67	71.07

Code Red LightGBM model: F-Score = 0
Model is unable to learn the data properties: highly unbalanced dataset

GBDT Models: Best Performance

- GBDT models offer shorter training time than recurrent neural networks and broad learning systems
- Models generated using the worm datasets exhibit higher accuracy than using ransomware datasets
- Best F-Scores:
  - CatBoost: Code Red dataset using 37 features: 81.30 %
  - XGBoost: WestRock dataset using 8 features: 72.96 %

Z. Li, A. L. Gonzalez Rios, and Lj. Trajkovic, "Machine learning for detecting anomalies and intrusions in communication networks," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 7, pp. 2254–2264, July 2021.

Discussion

- Increased number of BGP update announcements during worm attacks are more evident than ransomware attacks thus leading to better accuracy
- Worms (2001, 2003) and ransomware (2017, 2021) datasets have been collected decades apart and employ different attack mechanisms:
  - anomalous network activities were easier to observe in the early development of the Internet
  - Internet expansions, increased digital presence, device connectivity, and malicious activity have impacted traffic behavior and have made the detection of anomalous activities challenging
- While increased traffic volume was easily observed during the worm attacks, the distinction between regular and anomalous traffic during the ransomware attacks is less evident

Conclusions

- Extra-trees proved to be the best among various feature selection approaches to identify the important features
- K-S tests indicated that heavy-tailed distributions are a suitable fit for various BGP features
- Burr distribution was accepted for Code Red and WannaCrypt common features highlighting their underlying similarities
- Experimental results indicated that GBDT models offer a short training time desired for real-time intrusion detection systems
- Identifying anomalies based on F-Score in the case of the WannaCrypt ransomware attack remains a challenging task

References: Tools

- Python: https://pypi.org
  Pandas: https://pandas.pydata.org/
- PyTorch: https://pytorch.org/docs/stable/nn.html
- Google Colab: https://colab.research.google.com/
- zebra-dump-parser:
  https://github.com/rfc1036/zebra-dump-parser
- BGP C# tool:
  http://www.sfu.ca/~ljilja/cnl/projects/BGP_datasets/index.html
- IEEE DataPort
  Border Gateway Protocol (BGP) datasets:
  - https://ieee-dataport.org/open-access/border-gateway-protocol-bgp-routing-records-reseaux-ip-europeens-ripe-and-bcnet
  - https://ieee-dataport.org/open-access/border-gateway-protocol-bgp-routing-records-route-views

Publications: http://www.sfu.ca/~ljilja

Journal publication:

- Z. Li, A. L. Gonzalez Rios, and Lj. Trajkovic, “Machine learning for detecting the WestRock ransomware attack using BGP routing records,” *IEEE Communications* Magazine, vol. 61, no. 3, pp. 20–26, Mar. 2023.
- Z. Li, A. L. Gonzalez Rios, and Lj. Trajkovic, “Machine learning for detecting anomalies and intrusions in communication networks,” *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 7, pp. 2254-2264, July 2021.

Book chapters:

- Q. Ding, Z. Li, S. Haeri, and Lj. Trajković, “Application of machine learning techniques to detecting anomalies in communication networks: datasets and feature selection algorithms” in *Cyber Threat Intelligence,* M. Conti, A. Dehghantanha, and T. Dargahi, Eds., Berlin: Springer, pp. 47–70, 2018.
- Z. Li, Q. Ding, S. Haeri, and Lj. Trajković, “Application of machine learning techniques to detecting anomalies in communication networks: classification algorithms” in *Cyber Threat Intelligence,* M. Conti, A. Dehghantanha, and T. Dargahi, Eds., Berlin: Springer, pp. 71–92, 2018.

Publications: http://www.sfu.ca/~ljilja

Conference publications:

- Z. Li and Lj. Trajković, “CyberDefense: tool for detecting network anomalies and intrusions,” *IEEE Int. Conf. Syst., Man, Cybern.*, Maui, HI, USA, Oct. 2023, to be presented.
- T. Sharma, K. Patni, Z. Li, and Lj. Trajković, “Deep echo state networks for detecting Internet worm and ransomware attacks” *IEEE Int. Symp. Circuits and Systems*, Monterey, CA, USA, May 2023.
- Z. Li, A. L. Gonzalez Rios, and Lj. Trajković, “Classifying denial of service attacks using fast machine learning algorithms,” in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Melbourne, Australia, Oct. 2021, pp. 1221-1226 (virtual).
- K. Bekshentayeva and Lj. Trajkovic, "Detection of denial of service attacks using echo state networks," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Melbourne, Australia, Oct. 2021, pp. 1227-1232 (virtual).
- Z. Li, A. L. Gonzalez Rios, and Lj. Trajković, “Detecting Internet worms, ransomware, and blackouts using recurrent neural networks,” in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Toronto, Canada, Oct. 2020, pp. 2165-2172 (virtual).
- A. L. Gonzalez Rios, Z. Li, K. Bekshentayeva, and Lj. Trajković, “Detection of denial of service attacks in communication networks,” in *Proc. IEEE Int. Symp. Circuits and Systems*, Seville, Spain, Oct. 2020 (virtual).


Thank you!