Detecting Internet Worms, Ransomware, and Blackouts Using Recurrent Neural Networks

Zhida Li, Ana Laura Gonzalez Rios, and Ljiljana Trajković

Communication Networks Laboratory

http://www.ensc.sfu.ca/cnl

School of Engineering Science

Simon Fraser University, Vancouver, British Columbia

Canada

Roadmap

- Introduction
- BGP data collections: RIPE, Route Views
- BGP anomalies:
  Slammer, WannaCrypt, Moscow blackout
- BGP datasets
- Experimental procedure
  - Deep learning: multi-layer networks
  - BGP anomaly detection
- Performance comparison: LSTM, GRU
- Conclusion and references

Roadmap

- Introduction
- BGP data collections: RIPE, Route Views
- BGP anomalies:
  Slammer, WannaCrypt, Moscow blackout
- BGP datasets
- Experimental procedure
  - Deep learning: multi-layer networks
  - BGP anomaly detection
- Performance comparison: LSTM, GRU
- Conclusion and references

Border Gateway Protocol

- BGP's main function is to optimally route data between Autonomous Systems
- Types of BGP messages:
  - open, keepalive, update, and notification
- BGP anomalies:
  - worms, ransomware attacks, routing misconfigurations, Internet Protocol prefix hijacks, and link failures
- Collections of BGP update messages:
  - Réseaux IP Européens (RIPE)
  - Route Views

Machine Learning Algorithms

- Supervised machine learning algorithms:
  - Support vector machine: SVM
  - Broad learning system: BLS
  - Long short-term memory: LSTM
  - Gated recurrent unit: GRU

Roadmap

- Introduction
- BGP data collections: RIPE, Route Views
- BGP anomalies:
  Slammer, WannaCrypt, Moscow blackout
- BGP datasets
- Experimental procedure
  - Deep learning: multi-layer networks
  - BGP anomaly detection
- Performance comparison: LSTM, GRU
- Conclusion and references

RIPE and Route Views

- RIPE:
  - RIPE Network Coordination Centre project established in 2001 to collect and store routing data from several ASes worldwide
  - Remote route collectors installed at major topologically interesting Internet points for collection of BGP data
- Route Views:
  - University of Oregon project to collect real-time BGP routing data from various backbone routers and locations worldwide

Roadmap

- Introduction
- BGP data collections: RIPE, Route Views
- BGP anomalies:
  Slammer, WannaCrypt, Moscow blackout
- BGP datasets
- Experimental procedure
  - Deep learning: multi-layer networks
  - BGP anomaly detection
- Performance comparison: LSTM, GRU
- Conclusion and references

BGP Anomalies

- Slammer:
  - The fastest worm that self-propagated by using the User Datagram Protocol
  - Infected Microsoft SQL servers through a small piece of code that generated IP addresses at random
- WannaCrypt:
  - Data files are encrypted
  - Ransom is requested
- Moscow blackout:
  - Caused a complete shutdown of the Chagino substation of the Moscow energy ring
  - Caused the failure of the Internet traffic exchange

Roadmap

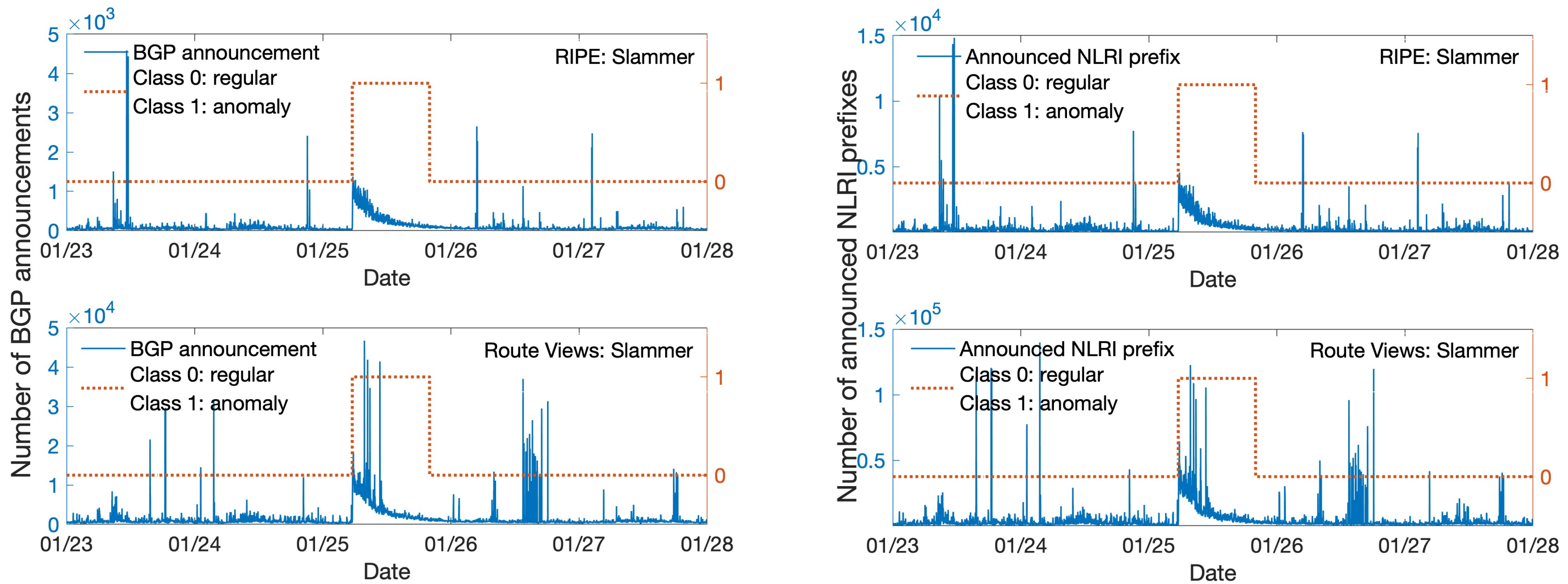
- Introduction
- BGP data collections: RIPE, Route Views
- BGP anomalies:
  Slammer, WannaCrypt, Moscow blackout
- BGP datasets
- Experimental procedure
  - Deep learning: multi-layer networks
  - BGP anomaly detection
- Performance comparison: LSTM, GRU
- Conclusion and references

BGP Datasets

- Anomalous data: days of the attack
- Regular data: two days prior and two days after the attack
- 37 numerical features from BGP update messages
- Training and test datasets are created based on the percentages of anomalous data:
  - training: 60%
  - testing: 40%

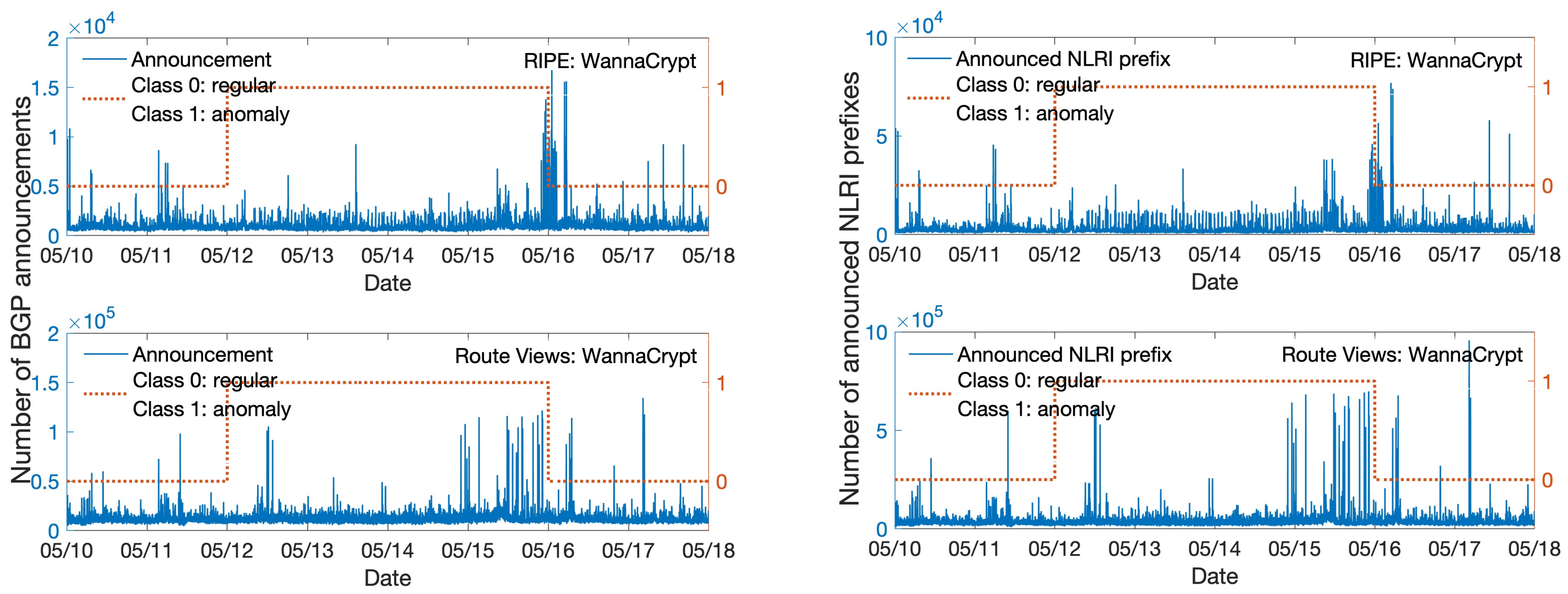
BGP Dataset: Slammer

- BGP announcements and announced NLRI prefixes:



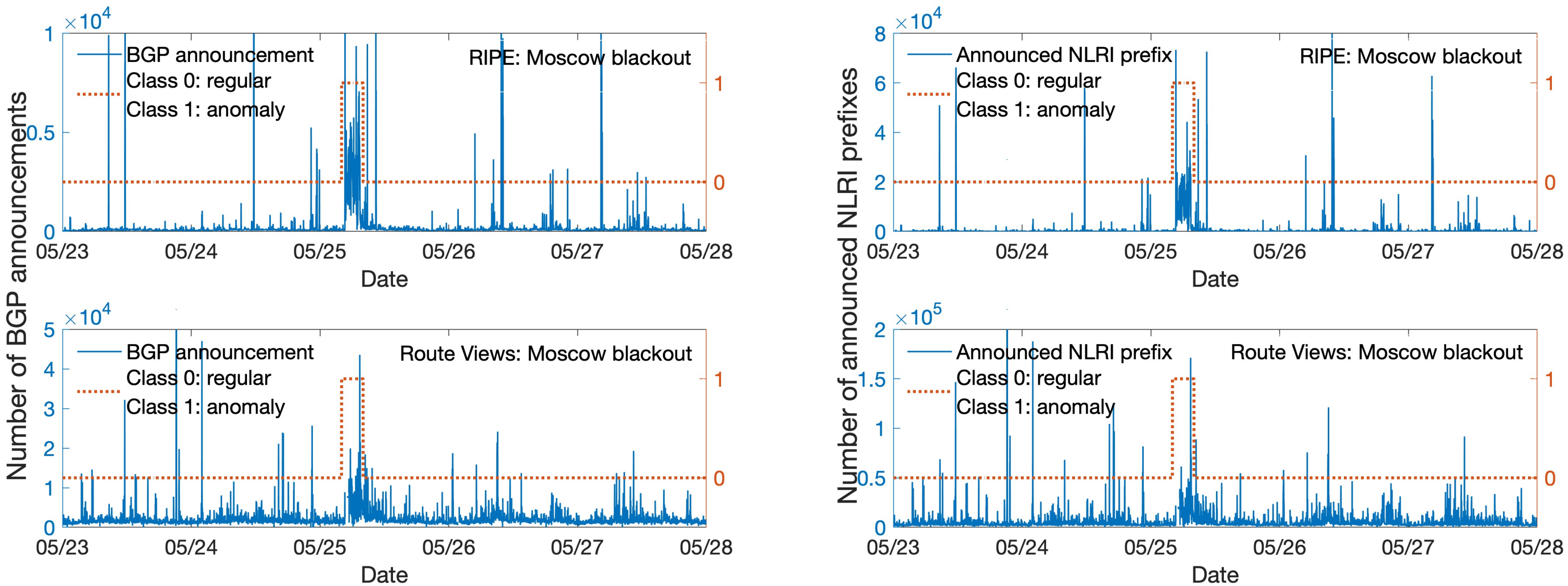
BGP Dataset: WannaCrypt

- BGP announcements and announced NLRI prefixes:



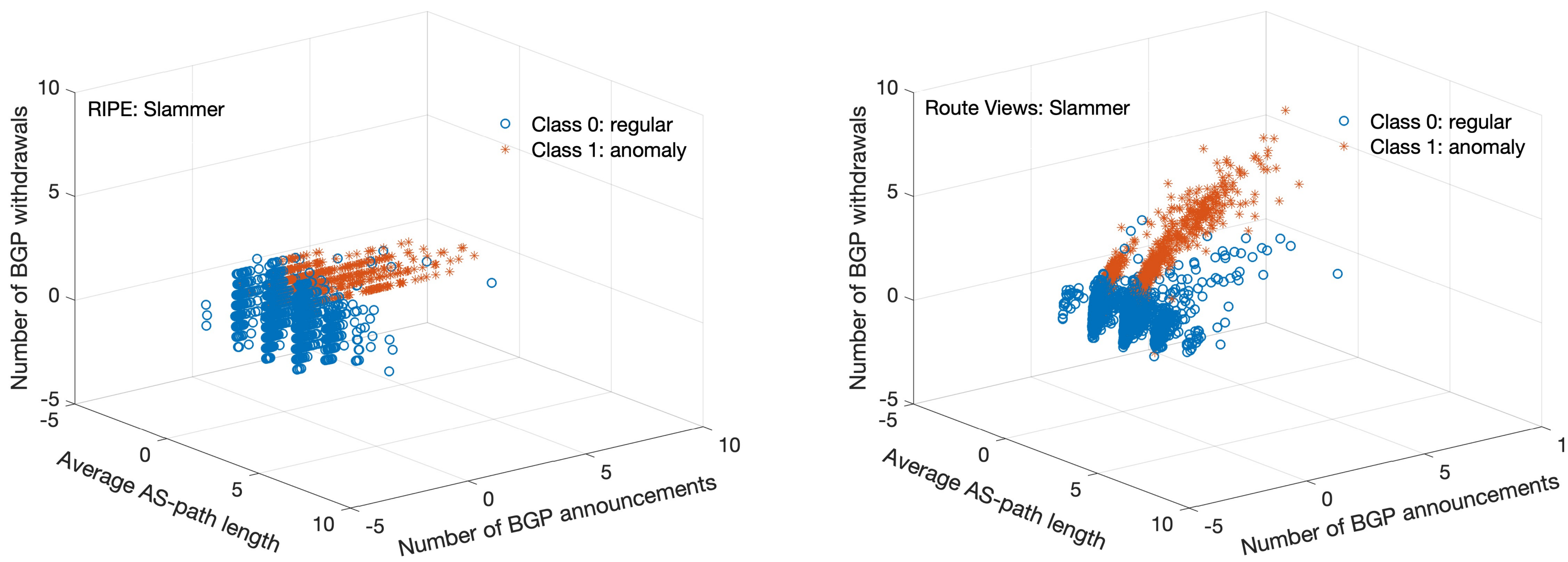
BGP Dataset: Moscow Blackout

- BGP announcements and announced NLRI prefixes:



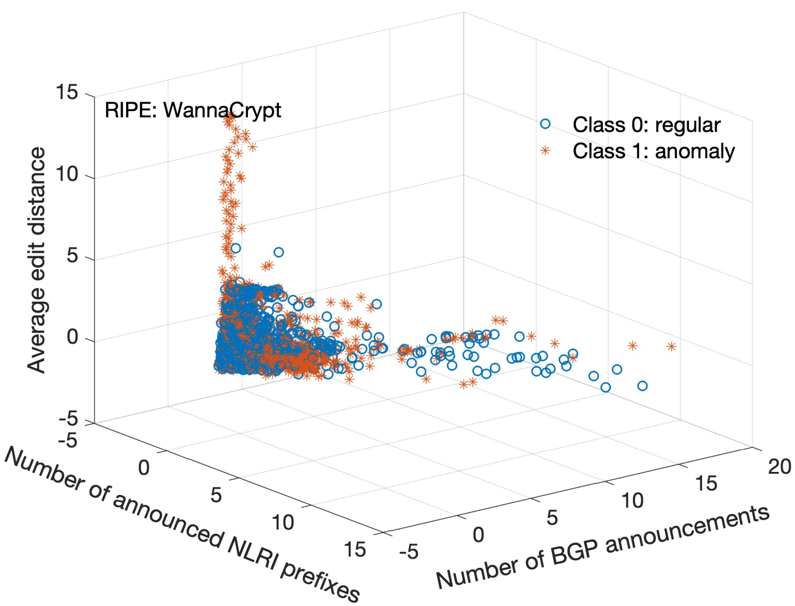
BGP Dataset: Slammer

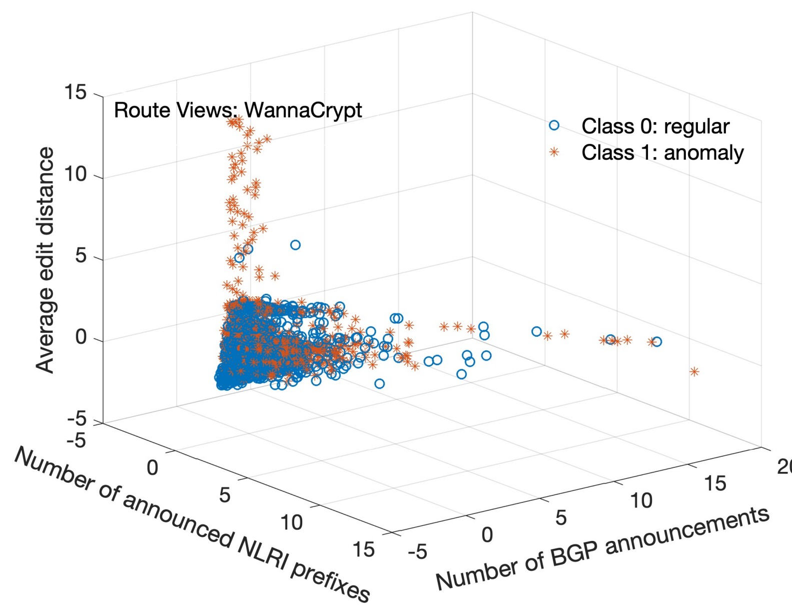
- Average AS-path length vs. number of BGP announcements vs. number of BGP withdrawals:



BGP Dataset: WannaCrypt

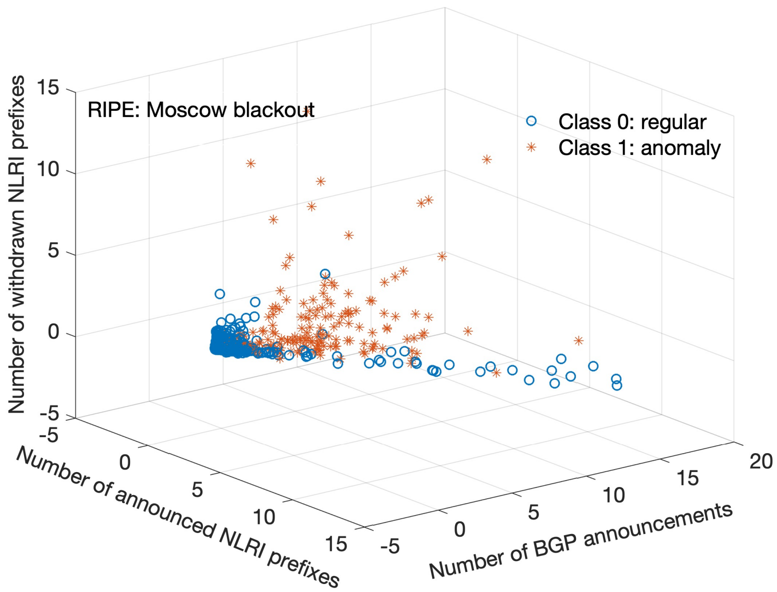
- Number of announced NLRI prefixes vs. number of BGP announcements vs. average edit distance:

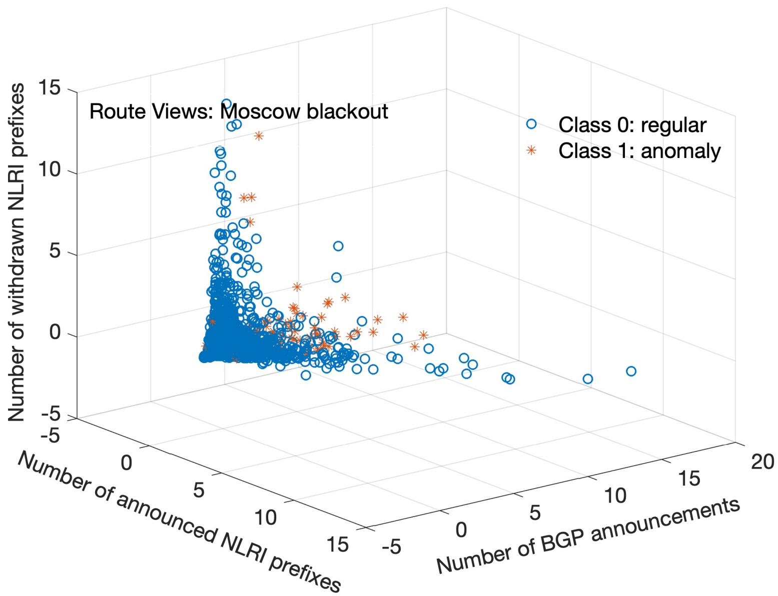




BGP Dataset: Moscow Blackout

- Number of announced NLRI prefixes vs. number of BGP announcements vs. number of withdrawn NLRI prefixes:





BGP Datasets

- Duration of BGP events and number of data points

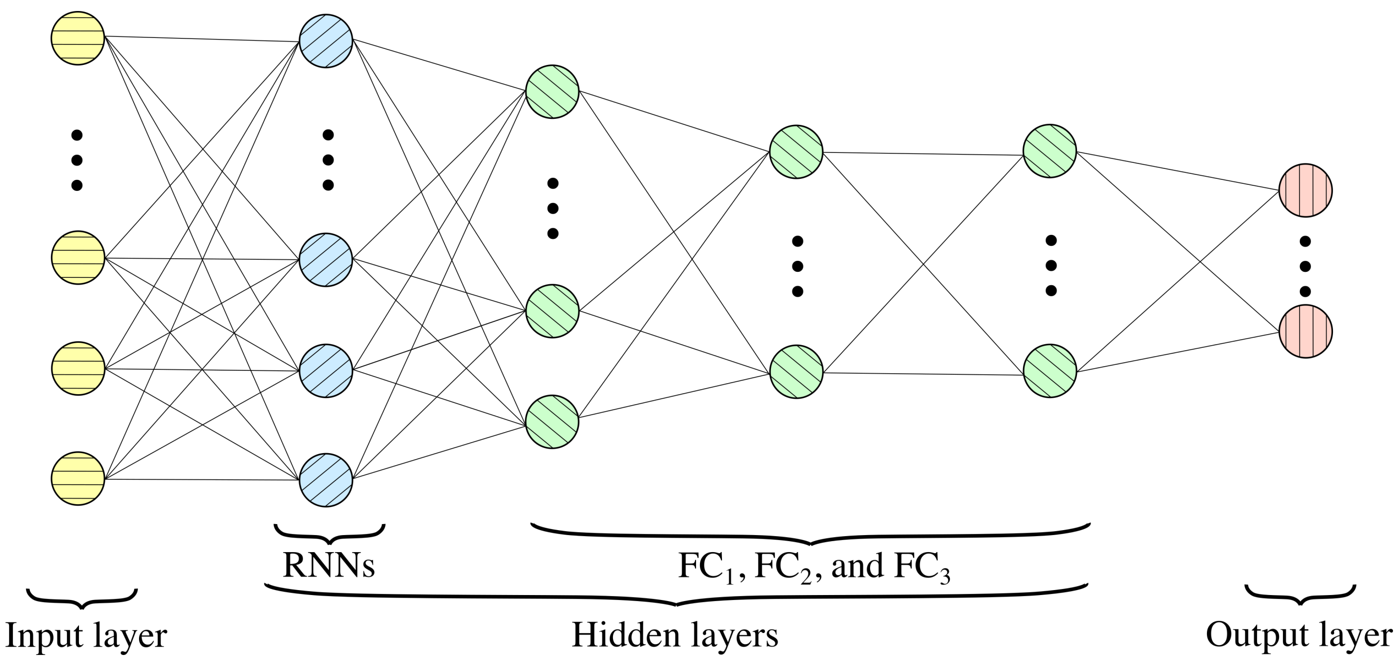
Collection site	Dataset	Regular (min)	Anomaly (min)	Regular (training)	Anomaly (training)	Regular (test)	Anomaly (test)	Collection date	
								Start	End
RIPE	Slammer	6,331	869	3,210	530	3,121	339	23.01.2003 00:00:00	27.01.2003 23:59:59
	WannaCrypt	5,760	5,760	2,880	3,420	2,880	2,340	10.05.2017 00:00:00	17.05.2017 23:59:59
	Moscow b/o	6,960	240	3,120	180	3,840	60	23.05.2005 00:00:00	27.05.2005 23:59:59
Route Views	Slammer	6,319	869	3,198	530	3,121	339	23.01.2003 00:00:00	27.01.2003 23:59:59
	WannaCrypt	5,760	5,760	2,880	3,420	2,880	2,340	10.05.2017 00:00:00	17.05.2017 23:59:59
	Moscow b/o	6,865	130	3,075	85	3,790	45	23.05.2005 00:00:00	27.05.2005 23:59:59

Roadmap

- Introduction
- BGP data collections: RIPE, Route Views
- BGP anomalies:
  Slammer, WannaCrypt, Moscow blackout
- BGP datasets
- Experimental procedure
  - Deep learning: multi-layer networks
  - BGP anomaly detection
- Performance comparison: LSTM, GRU
- Conclusion and references

Deep Learning: Multi-Layer Networks

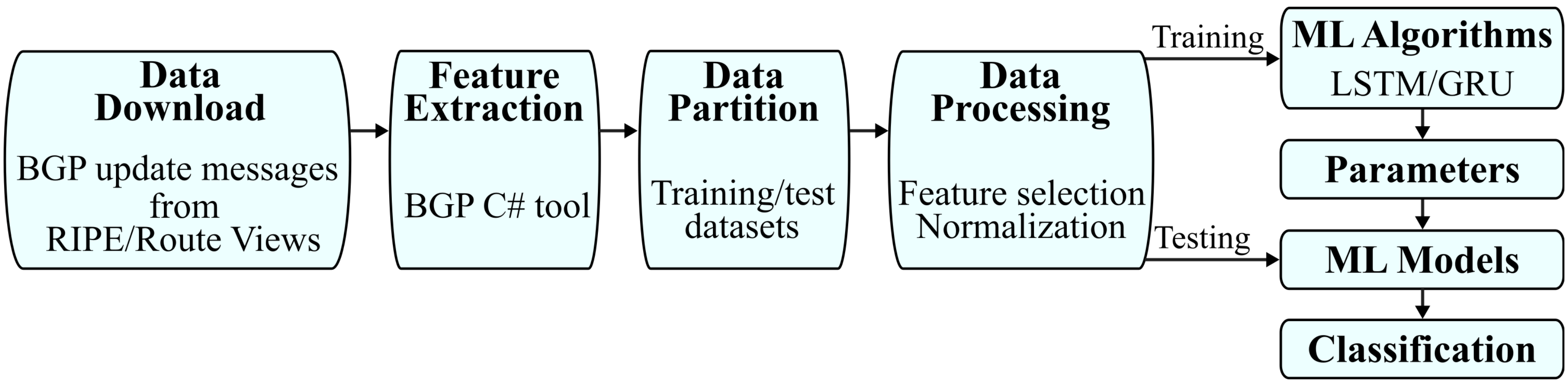
- 37 RNNs, 80 (Slammer)/64 (WannaCrypt)/64 (Moscow blackout) FC1, 32 FC2, and 16 FC3 fully connected (FC) hidden nodes



RNN Model Parameters

Parameter	Value	Best selection
Length of input sequence	5, 10, 20, 50, 100	Slammer: 10 WannaCrypt: 100 Moscow b/o: 100 (RIPE), 20 (Route Views)
Number of epochs	30, 50, 100	30
Number of hidden nodes	80, 64, 32, 16	Slammer: FC1 = 80, FC2 = 32, FC3 = 16 WannaCrypt/Moscow b/o: F11 = 64, FC2 = 32, FC3 = 16
Dropout rate	0.2, 0.4, 0.6	0.4
Learning rate	0.01, 0.1	0.01

BGP Anomaly Detection



Roadmap

- Introduction
- BGP data collections: RIPE, Route Views
- BGP anomalies:
  Slammer, WannaCrypt, Moscow blackout
- BGP datasets
- Experimental procedure
  - Deep learning: multi-layer networks
  - BGP anomaly detection
- Performance comparison: LSTM, GRU
- Conclusion and references

LSTM and GRU RNN Models

- We evaluate performance of LSTM and GRU models with various of hidden layers:
  - 2: LSTM2 and GRU2
  - 3: LSTM3 and GRU3
  - 4: LSTM4 and GRU4
- Performance evaluation is based on:
  - Accuracy
  - F-Score

LSTM RNN Models

Model	Dataset	Accuracy (%)		F-Score (%)	
		RIPE	Route Views	RIPE	Route Views
$LSTM_2$	Slammer	92.98	91.24	72.42	69.11
	WannaCrypt	58.08	67.23	61.48	70.14
	Moscow b/o	99.21	96.23	75.20	5.26
$LSTM_3$	Slammer	90.90	95.72	67.29	81.77
	WannaCrypt	65.48	64.35	63.22	67.16
	Moscow b/o	98.38	97.77	55.94	32.00
$LSTM_4$	Slammer	92.49	91.39	70.72	69.34
	WannaCrypt	57.94	72.29	62.42	73.86
	Moscow b/o	97.46	95.81	36.94	18.37

GRU RNN Models

Model	Dataset	Accuracy (%)		F-Score (%)	
		RIPE	Route Views	RIPE	Route Views
GRU_2	Slammer	91.88	92.60	69.42	72.59
	WannaCrypt	57.27	72.58	60.56	74.21
	Moscow b/o	97.64	98.30	41.77	32.99
GRU_3	Slammer	91.76	93.24	68.72	74.34
	WannaCrypt	52.85	72.63	53.96	74.14
	Moscow b/o	98.38	97.51	57.14	28.57
GRU_4	Slammer	92.14	93.15	70.11	74.04
	WannaCrypt	52.15	68.71	52.70	71.61
	Moscow b/o	97.92	97.20	49.06	35.15

LSTM and GRU RNN Models: Observations

- Increasing the number of the hidden layers in $LSTM_4$ model may have resulted in over-fitting
- The best accuracy and F-Score generated by RNN models using Slammer and WannaCrypt data collected by Route Views are higher than data collected by RIPE
- Better classification results were achieved using Moscow blackout data collected by RIPE being more reliable than Route Views data

Roadmap

- Introduction
- BGP data collections: RIPE, Route Views
- BGP anomalies:
  Slammer, WannaCrypt, Moscow blackout
- BGP datasets
- Experimental procedure
  - Deep learning: multi-layer networks
  - BGP anomaly detection
- Performance comparison: LSTM, GRU
- Conclusion and references

Conclusion

- BGP update messages collected by RIPE and Route Views data collection sites were used to classify Slammer, WannaCrypt, and Moscow blackout anomalous events
- RNN models with two and three hidden layers often exhibited the best performance
- BGP update messages collected by Route Views generated the best accuracy and F-Score for Slammer and WannaCrypt
- Classification models for Slammer dataset offered better results due to better spatial separation between regular and anomalous classes

Roadmap

- Introduction
- BGP data collections: RIPE, Route Views
- BGP anomalies:
  Slammer, WannaCrypt, Moscow blackout
- BGP datasets
- Experimental procedure
  - Deep learning: multi-layer networks
  - BGP anomaly detection
- Performance comparison: LSTM, GRU
- Conclusion and references

References: Data Sources and Tools

- RIPE NCC:
  https://www.ripe.net
- University of Oregon Route Views project:
  http://www.routeviews.org
- PyTorch
  https://pytorch.org/docs/stable/nn.html
- 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-routing-records-reseaux-ip-europeens-ripe-and-bcnet

References

Intrusion Detection:

- B. Al-Musawi, P. Branch, and G. Armitage, “BGP anomaly detection techniques: a survey,” *IEEE Commun. Surv. Tut.*, vol. 19, no. 1, pp. 377–396, 2017.
- Y. Song, A. Venkataramani, and L. Gao, “Identifying and addressing reachability and policy attacks in ‘secure’ BGP,” *IEEE/ACM Trans. Netw.*, vol. 24, no. 5, pp. 2969–2982, Oct. 2016.
- A. Lutu, M. Bagnulo, C. Pelsser, O. Maennel, and J. Cid-Sueiro, “The BGP visibility toolkit: detecting anomalous Internet routing behavior,” *IEEE/ACM Trans. Netw.*, vol. 24, no. 2, pp. 1237–1250, Apr. 2016.
- D. Dolev, S. Jamin, O. Mokryn, and Y. Shavitt, “Internet resiliency to attacks and failures under BGP policy routing,” *Comput. Netw.*, vol. 50, no. 16, pp. 3183–3196, Nov. 2006.

RNNs:

- K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, “LSTM: a search space odyssey,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017.
- K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning phrase representations using RNN encoder–decoder for statistical machine translations,” in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Doha, Qatar, Oct. 2014, pp. 1724–1734.

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

- 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.
- Z. Li, A. L. Gonzalez Rios, G. Xu, and Lj. Trajković, “Machine learning techniques for classifying network anomalies and intrusions,” in *Proc. IEEE Int. Symp. Circuits and Systems*, Sapporo, Japan, May 2019.
- A. L. Gonzalez Rios, Z. Li, G. Xu, A. Dias Alonso, and Lj. Trajković, “Detecting network anomalies and intrusions in communication networks,” in *Proc. 23rd IEEE International Conference on Intelligent Engineering Systems 2019*, Gödöllő, Hungary, Apr. 2019, pp. 29–34.
- Z. Li, P. Batta, and Lj. Trajković, “Comparison of machine learning algorithms for detection of network intrusions,” in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Miyazaki, Japan, Oct. 2018, pp. 4248–4253.
- P. Batta, M. Singh, Z. Li, Q. Ding, and Lj. Trajković, “Evaluation of support vector machine kernels for detecting network anomalies,” in *Proc. IEEE Int. Symp. Circuits and Systems*, Florence, Italy, May 2018, pp. 1–4.
- 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.

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

- 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.
- Q. Ding, Z. Li, P. Batta, and Lj. Trajković, "Detecting BGP anomalies using machine learning techniques," in *Proc. IEEE International Conference on Systems, Man, and Cybernetics (SMC 2016),* Budapest, Hungary, Oct. 2016, pp. 3352–3355.
- Y. Li, H. J. Xing, Q. Hua, X.-Z. Wang, P. Batta, S. Haeri, and Lj. Trajković, "Classification of BGP anomalies using decision trees and fuzzy rough sets," in *Proc. IEEE International Conference on Systems, Man, and Cybernetics*, *SMC 2014*, San Diego, CA, October 2014, pp. 1312–1317.
- N. Al-Rousan, S. Haeri, and Lj. Trajković, "Feature selection for classification of BGP anomalies using Bayesian models," in *Proc. International Conference on Machine Learning and Cybernetics, ICMLC 2012*, Xi'an, China, July 2012, pp. 140–147.
- N. Al-Rousan and Lj. Trajković, "Machine learning models for classification of BGP anomalies," in *Proc. IEEE Conf. High Performance Switching and Routing*, *HPSR 2012*, Belgrade, Serbia, June 2012, pp. 103–108.