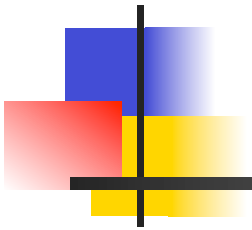
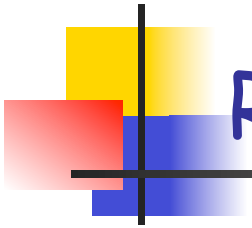


Performance Evaluation of BGP Anomaly Classifiers

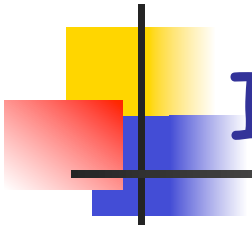
Marijana Ćosović and Slobodan Obradović
{marijana.cosovic, slobo.obradovic}@gmail.com
University of East Sarajevo
East Sarajevo, Bosnia and Herzegovina

Ljiljana Trajković
ljilja@cs.sfu.ca
Simon Fraser University, Vancouver, British Columbia
Canada



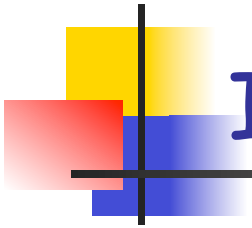
Roadmap

- Introduction
- BGP data
- Waikato Environment for Knowledge Analysis (Weka)
- Data transformation
- Performance measures
- Classification models
- Conclusion
- References



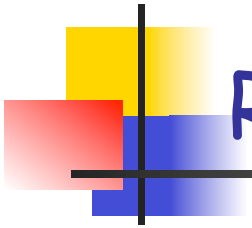
Introduction

- Internet worm attacks may induce routing information updates
- Network reachability information is contained in Border Gateway Protocol (BGP) update messages and stored in Routing Information Base (RIB)
- BGP anomaly detection systems employ machine learning techniques to mine network data:
  - Meta learning, a subfield of machine learning, deals with automatic detection of data models



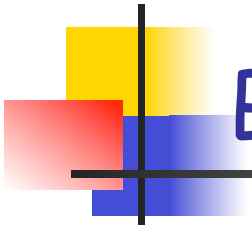
Introduction

- Internet routing anomalies can be detected from BGP update messages
- We consider:
  - Three data sets of known Internet anomalies
  - Set of 15 features based on BGP update messages
- Anomaly classifiers:
  - Naïve Bayes
  - Decision Tree
  - Support Vector Machine
- We use measure performance indices to compare classifiers



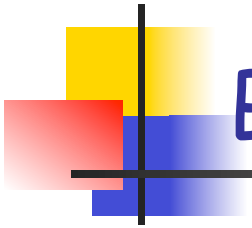
Roadmap

- Introduction
- BGP data
- Waikato Environment for Knowledge Analysis (Weka)
- Data transformation
- Performance measures
- Classification models
- Conclusion
- References



BGP Data

- Routing Information Services project began in 2001 by the Réseaux IP Européens Network Coordination Centre to collect and store Internet routing data:
  - BGP update messages are collected by Remote Route Collectors
  - Format of BGP update messages is multi-threaded routing toolkit
  - Data are collected during the periods of the Internet anomalies: Slammer, Nimda, Code Red I



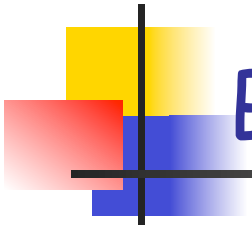
BGP Data: Slammer and Nimda

- Slammer
  - Microsoft SQL servers were infected through a small piece of code that generated IP addresses at random
  - Attack duration: 16h
- Nimda
  - Exploited vulnerabilities in the Internet Information Services web servers for the Internet Explorer 5
  - Attack duration: 59h

BGP Data: Code Red I

- Code Red I:
  - Replicated itself by exploiting weakness of the IIS servers and searched for vulnerable servers
  - Attack duration: 10h

	Number of events		Number of features	Number of classes
	Anomaly	Regular		
Slammer	869	6,331	15	2
Nimda	3,521	3,679	15	2
Code Red I	600	6,600	15	2



BGP Data: Feature Extraction

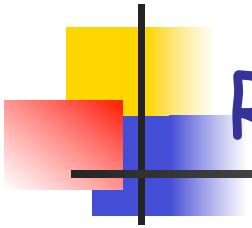
- Sample BGP messages: every minute during a five-day interval
- 7,200 samples for each of the three anomalous events
- Produced imbalanced datasets
- 15 features were extracted from BGP update messages
- Classified as volume and AS-Path features

BGP Data: Feature Extraction

Feature	Name	Category
1	Number of announcements	Volume
2	Number of withdrawals	Volume
3	Number of announced Network Layer Reachability Information (NLRI) prefixes	Volume
4	Number of withdrawn Network Layer Reachability Information (NLRI) prefixes	Volume
5	Average AS-PATH length	AS-Path
6	Maximum AS-PATH length	AS-Path
7	Average unique AS-PATH length	AS-Path
8	Number of duplicate announcements	Volume

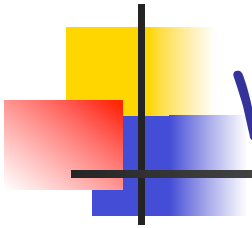
BGP Data: Feature Extraction

Feature	Name	Category
9	Number of duplicate withdrawals	Volume
10	Number of implicit withdrawals	Volume
11	Average edit distance	AS-Path
12	Maximum edit distance	AS-Path
13	Number of Exterior Gateway Protocol (EGP) packets	Volume
14	Number of Interior Gateway Protocol (IGP) packets	Volume
15	Number of incomplete packets	Volume



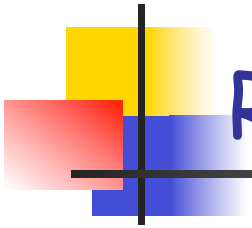
Roadmap

- Introduction
- BGP data
- Waikato Environment for Knowledge Analysis (Weka)
- Data transformation
- Performance measures
- Classification models
- Conclusion
- References



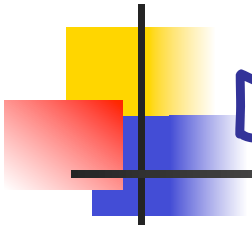
Weka

- Waikato Environment for Knowledge Analysis (Weka):
  - Open source software tool distributed under the GNU General Public License
  - Framework for implementation of machine learning algorithms
  - Version 3.7.11
- Weka classifier implementations:
  - Naïve Bayes
  - Decision Tree (C4.5 algorithm)
  - Support Vector Machine (wrapper to the LibSVM)



Roadmap

- Introduction
- BGP data
- Waikato Environment for Knowledge Analysis (Weka)
- Data transformation
- Performance measures
- Classification models
- Conclusion
- References



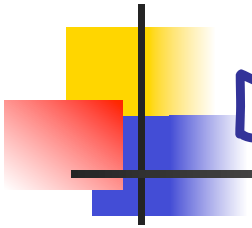
Data Transformations

Techniques used to make input data more responsive to machine learning algorithms:

- Feature discretization:
  - The minimum description length (MDL) discretization method is the most commonly used supervised discretization algorithm (implemented in Weka)
  - Benefits: clarity, better results, and faster convergence

Weka FilteredClassifier tool:

- Discretization of intervals obtained from training data

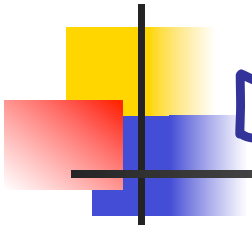


Data Transformations

- Feature selection:
  - Performed to minimize the number of features that a machine learning algorithm should consider by disregarding redundant or unrelated features
  - Benefits: reduces dimensionality, computational complexity, and memory usage

Weka Attribute Selection tool:

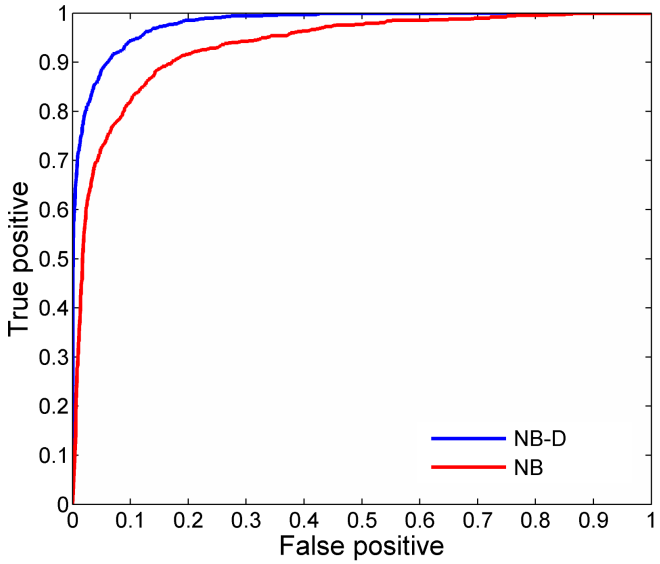
- Feature subsets evaluated using Attribute subset evaluator
- Search method enables search of all possible subsets

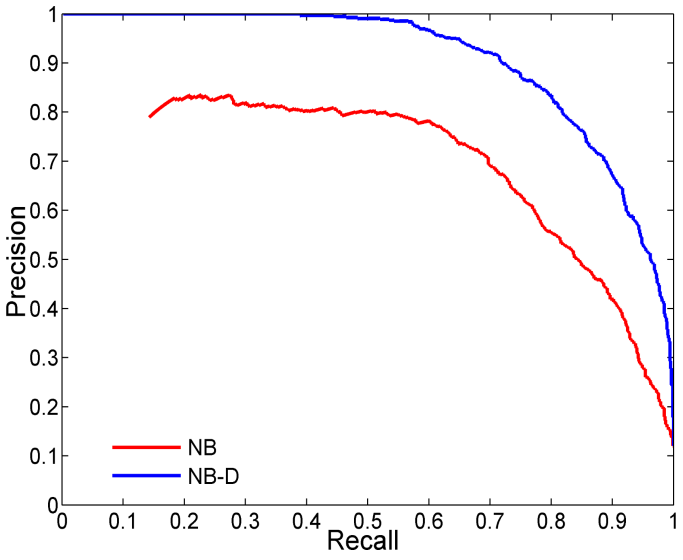


Data Transformations

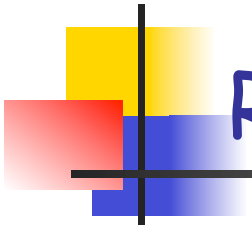
- Receiver operating characteristics (ROCs):
  - Plots of True Positive Rate (TPR) as a function of False Positive Rate (FPR), for various parameters of a machine learning model
  - May be misleading if the number of positive and negative instances greatly differ
- Precision-recall (PR) curves: used in machine learning tasks in case of class imbalance

Data Transformations: ROC and PR





- The ROC (left) and PR (right) curves for the NB classifier of Slammer anomaly with (NB-D) and without discretization (NB)
- Feature discretization improves ROC and PR curves



Roadmap

- Introduction
- BGP data
- Weka
- Data transformation
- Performance measures
- Classification models
- Conclusion
- References

Performance Measures

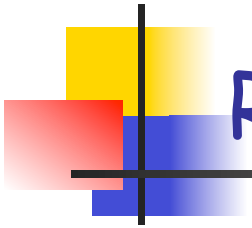
- We consider: Recall, Precision, F-measure, and the Matthews correlation coefficient (MCC)
  - Recall: ratio of identified anomalies (TP) and all labeled anomalies (true)
  - Precision: ratio of identified anomalies (TP) and all data points identified as anomalous
  - F-measure: harmonic mean of recall and precision
  - MCC: used for binary classification

$$F-measure = 2 \times \frac{recall \times precision}{recall + precision} \qquad MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

Performance Measures

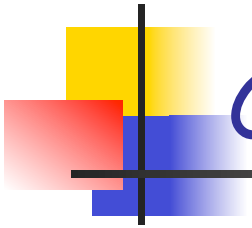
- TP: number of anomalous training data points classified as anomaly
- FP: number of regular training data points classified as anomaly
- FN: number of anomalous training data points classified as regular
- TN: number of regular training data points classified as regular

		Predicted class	
		Anomaly	Regular
Actual class	Anomaly (True)	TP	FN
	Regular (False)	FP	TN



Roadmap

- Introduction
- BGP data
- Waikato Environment for Knowledge Analysis (Weka)
- Data transformation
- Performance measures
- Classification models
- Conclusion
- References



Classification Models

- Weka FilteredClassifier used to evaluate supervised discretization:
  - NB-1, J48-1, and SVM-1
- Weka Threshold Selector used to optimize F-measure evaluation metrics:
  - NB-2, J48-2, and SVM-2
- Weka AttributeSelectedClassifier used to specify feature selection method and a learning algorithm as a part of the classification scheme:

Classification Models

- CfsSubsetEval and GreedyStepWise search methods: NB-3, J48-3, and SVM-3
- GainRatioAttributeEval and Ranker search methods: NB-4, J48-4, and SVM-4

- Weka wrapper feature selection methods used within AttributeSelectedClassifier to evaluate sets of relevant features:
  - ClassifierSubsetEval: NB-5, J48-5, and SVM-5
  - WrapperSubsetEval: NB-6, J48-6, and SVM-6

Classification Models: NB, J48, SVM

Classifiers	Methods	Models	Description
NB J48 SVM	Filter	1	Classifier trained on discretized data sets
		2	Classifier trained on data sets with F-measure optimized
		3	Correlation based feature subset evaluator (CfsSubsetEval) with Greedy Stepwise search method
		4	Gain ratio based feature evaluator with ranker for individual features
		5	Classifier subset evaluator using a classifier as a parameter for evaluation of sets of features on training data
	Wrapper	6	Wrapper subset evaluator using 5-folds cross-validation internally to estimate the accuracy of the learning scheme for a set of features

Classification Models: NB

Two tests were performed for each classifier on three different data sets:

- NB-2 model shows improvements over NB-1 model in all performance measures for Slammer, Nimda, and Code Red I data sets:

Data set	Model	F-measure	MCC	ROC	PR
Slammer	NB-1	0.767	0.741	0.980	0.907
	NB-2	0.807	0.781	0.980	0.906
Nimda	NB-1	0.745	0.493	0.826	0.817
	NB-2	0.758	0.483	0.826	0.816
Code Red I	NB-1	0.541	0.509	0.900	0.600
	NB-2	0.585	0.548	0.900	0.596

Classification Models: J48

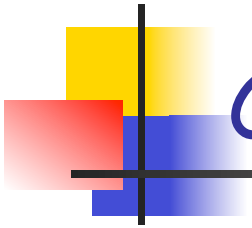
- J48-1 model performs better than J48-2 model in all performance measures for Slammer, Nimda and Code Red I data sets:

Data set	Model	F-measure	MCC	ROC	PR
Slammer	J48-1	0.844	0.825	0.967	0.879
	J48-2	0.826	0.802	0.966	0.876
Nimda	J48-1	0.755	0.518	0.815	0.774
	J48-2	0.753	0.485	0.814	0.773
Code Red I	J48-1	0.628	0.608	0.866	0.562
	J48-2	0.626	0.594	0.871	0.560

Classification Models: SVM

- SVM-1 classifier performs better on Slammer data set while SVM-2 performs better on Nimda and Code Red I data sets:

Data set	Model	F-measure	MCC	ROC	PR
Slammer	SVM-1	0.862	0.845	0.906	0.765
	SVM-2	0.855	0.837	0.980	0.926
Nimda	SVM-1	0.762	0.526	0.763	0.690
	SVM-2	0.767	0.506	0.844	0.825
Code Red I	SVM-1	0.564	0.542	0.729	0.372
	SVM-2	0.618	0.584	0.804	0.559



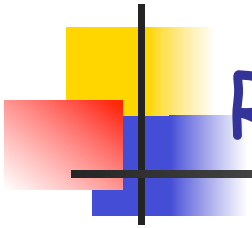
Classification Models

Additional four tests were performed for each classifier on three different data sets:

- Slammer and Code Red I data sets: wrapper methods for feature selection provide better results than filter methods
  - SVM-6 classifier model achieves the highest performance measures
  - Number of selected features reduced to eight
- Nimda data set: wrapper methods for feature selection slightly outperform filter methods

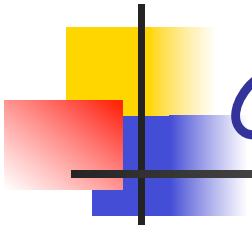
Classification Models: SVM

Data set	Model	F-measure	MCC	ROC	PR
Slammer	SVM-3	0.828	0.812	0.873	0.721
	SVM-4	0.856	0.843	0.886	0.766
	SVM-5	0.878	0.866	0.904	0.799
	SVM-6	0.880	0.867	0.907	0.800
Nimda	SVM-3	0.667	0.444	0.713	0.660
	SVM-4	0.684	0.423	0.709	0.646
	SVM-5	0.660	0.434	0.708	0.655
	SVM-6	0.661	0.435	0.709	0.655
Code Red I	SVM-3	0.568	0.566	0.716	0.396
	SVM-4	0.622	0.629	0.738	0.465
	SVM-5	0.623	0.624	0.743	0.459
	SVM-6	0.632	0.631	0.748	0.468



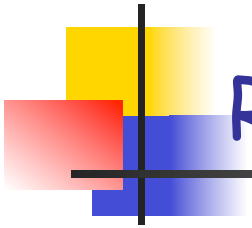
Roadmap

- Introduction
- BGP data
- Waikato Environment for Knowledge Analysis (Weka)
- Data transformation
- Performance measures
- Classification models
- Conclusion
- References



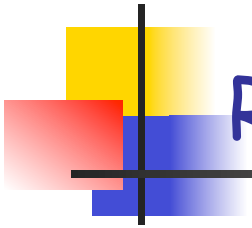
Conclusion

- Feature selection and classification algorithms were used to detect BGP anomalies
- We investigated performance measures of several BGP detection models
- Performance of classifiers depends on the employed data set
- No single classifier performs the best across all given data sets



References

- (2015, Jan.) Weka 3: Data Mining Software in Java [Online]. Available: http://www.cs.waikato.ac.nz/ml/weka/index.html.
- I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools* and Techniques, 3rd ed. San Francisco, CA, USA: Morgan Kaufmann, 2011, pp. 314–322.
- J. Dougherty, R. Kohavi, and M. Sahami, "Supervised and unsupervised discretization of continuous features," in *Proc. 12th Int. Conf. Mach. Learning*, Tahoe City, CA, USA, July 1995, pp. 194–202.



References

- 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 Int. Conf. Syst., Man, Cybern.*, San Diego, CA, USA, Oct. 2014, pp. 1331–1336.
- N. Al-Rousan, S. Haeri, and Lj. Trajković, "Feature selection for classification of BGP anomalies using Bayes models," in *Proc. Int. Conf. Mach. Learning Cybern.*, Xi'an, China, July 2012, pp. 140–147.
- N. Al-Rousan and Lj. Trajković, "Machine learning models for classification of BGP anomalies," in *Proc. 13th IEEE Int. Conf. High Performance Switching and Routing*, Belgrade, Serbia, June 2012, pp. 103–108.

References

- M. Ćosović, S. Obradović, and Lj. Trajković, "Using databases for a BGP data analysis," in *Proc. Int. Scientific Conf. UNITECH*, Gabrovo, Bulgaria, Nov. 2014, no. 2, pp. 367–370.
- M. Ćosović, S. Obradović, and Lj. Trajković, "Feature selection techniques for machine learning," in *Proc. Int. Scientific Conf. UNITECH*, Gabrovo, Bulgaria, Nov. 2013, no. 1, pp. 85–89.
- M. Ćosović, S. Obradović and Lj. Trajković, "Algorithms for investigation of abnormal BGP events," in *Proc. Int. Scientific Conf. UNITECH*, Gabrovo, Bulgaria, Nov. 2013, no. 2, pp. 253–257.