Communication Networks: Traffic Data, Network Topologies, and Routing Anomalies

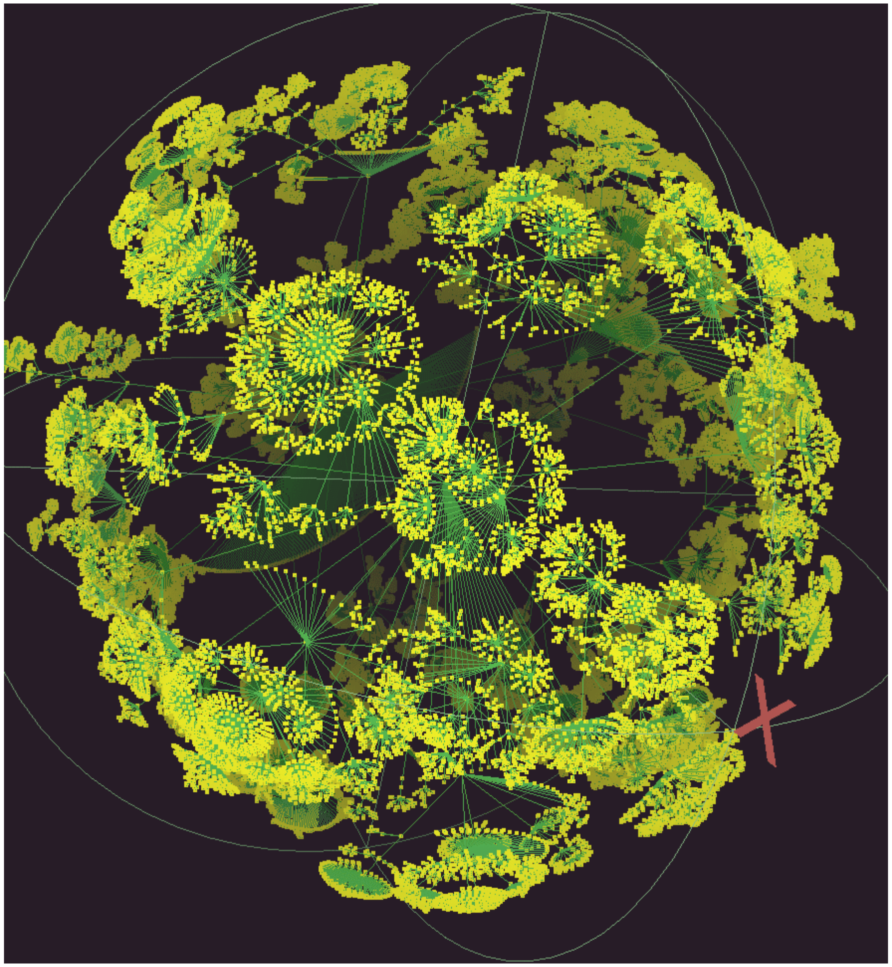
Ljiljana Trajković
ljilja@cs.sfu.ca

Communication Networks Laboratory
http://www.ensc.sfu.ca/cnl
School of Engineering Science
Simon Fraser University, Vancouver, British Columbia
Canada

Roadmap

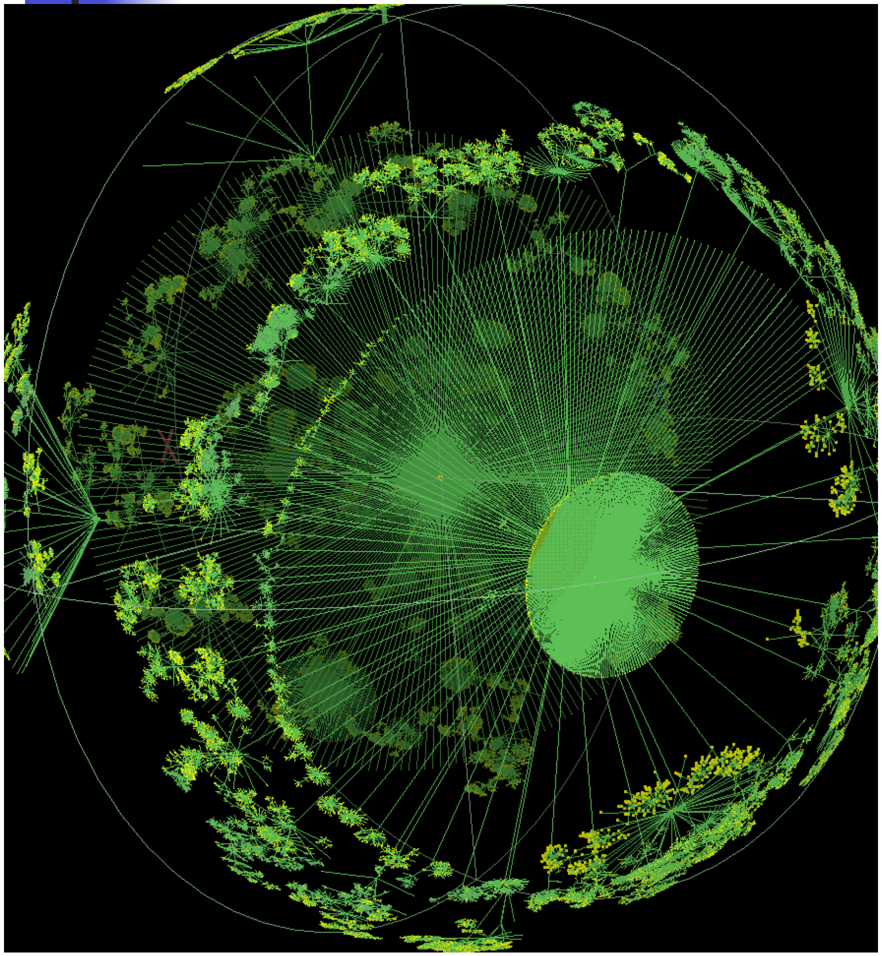
- Introduction
- Traffic collection, characterization, and modeling
- Case study: Collection of BCNET traffic
- Internet topology and spectral analysis of Internet graphs
- Machine learning models for feature selection and classification of traffic anomalies
- Conclusions

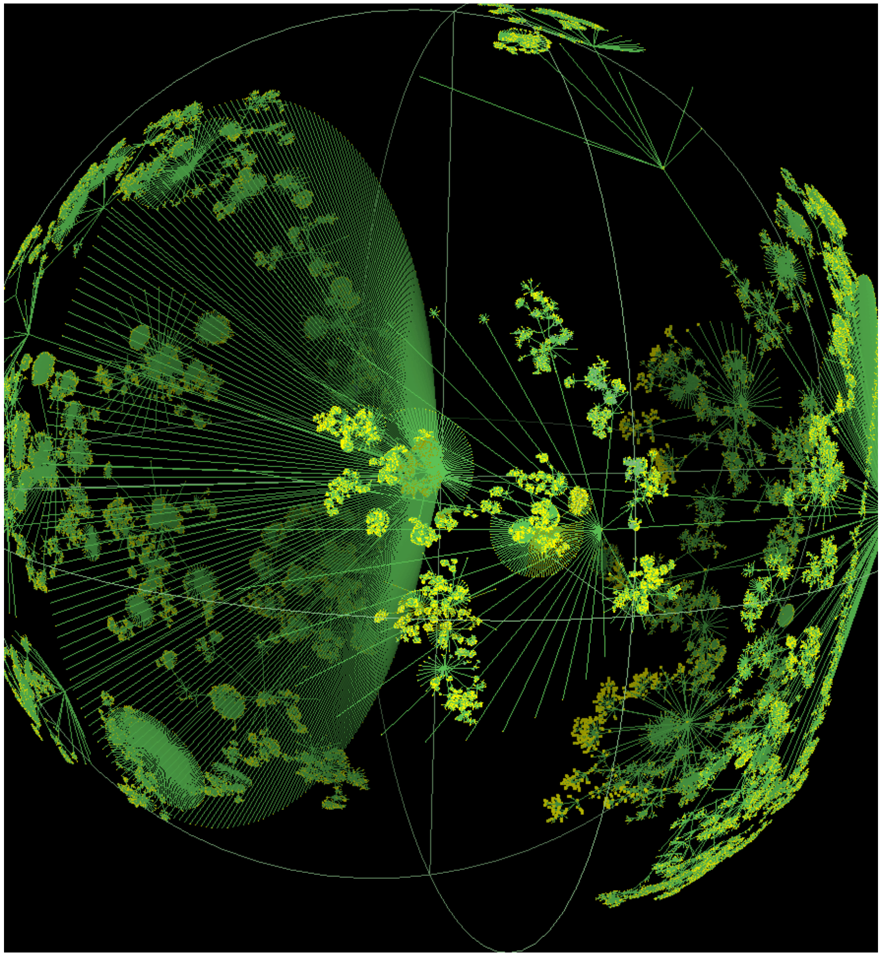
lhr: 535,102 nodes and 601,678 links



http://www.caida.org/home/

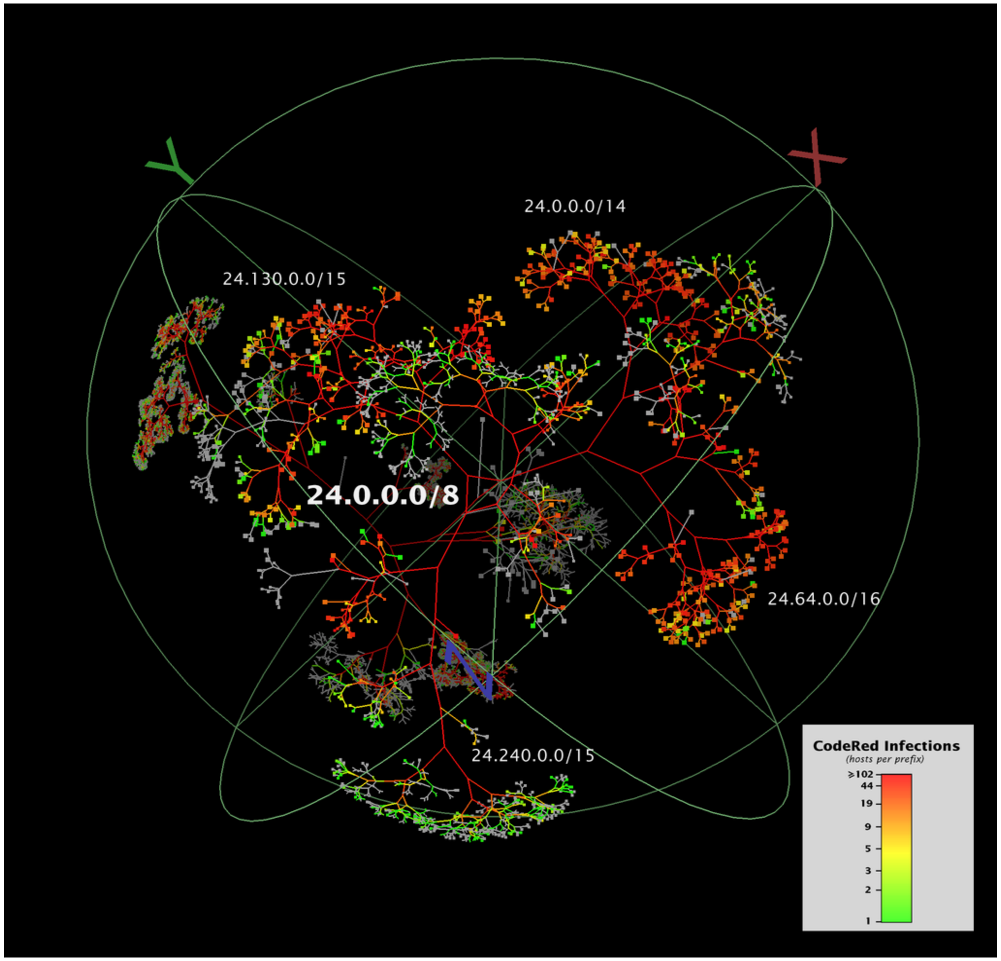
lhr: 535,102 nodes and 601,678 links





http://www.caida.org/home/

Code Red infection



http://www.caida.org/home/

Roadmap

- Introduction
- Traffic collection, characterization, and modeling
- Case study: Collection of BCNET traffic
- Internet topology and spectral analysis of Internet graphs
- Machine learning models for feature selection and classification of traffic anomalies
- Conclusions

Measurements of network traffic

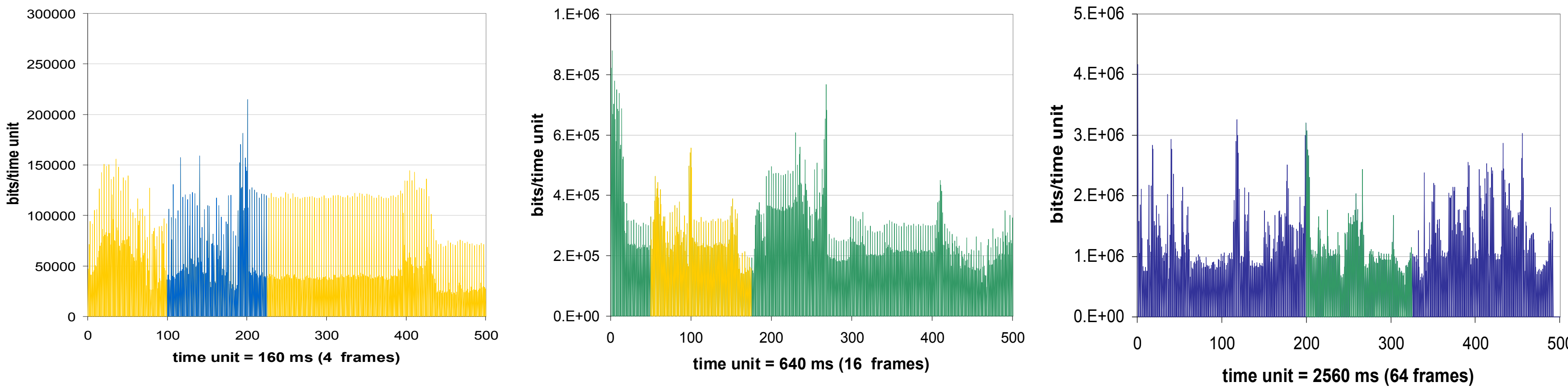
- Traffic measurements:
  - help understand characteristics of network traffic
  - are basis for developing traffic models
  - are used to evaluate performance of protocols and applications
- Traffic analysis:
  - provides information about the network usage
  - helps understand the behavior of network users
- Traffic prediction:
  - important to assess future network capacity requirements
  - used to plan future network developments

Traffic modeling: self-similarity

- Self-similarity implies a "fractal-like" behavior
- Data on various time scales have similar patterns
- Implications:
  - no natural length of bursts
  - bursts exist across many time scales
  - traffic does not become "smoother" when aggregated
  - it is unlike Poisson traffic used to model traffic in telephone networks
  - as the traffic volume increases, the traffic becomes more bursty and more self-similar

Self-similarity: influence of time-scales

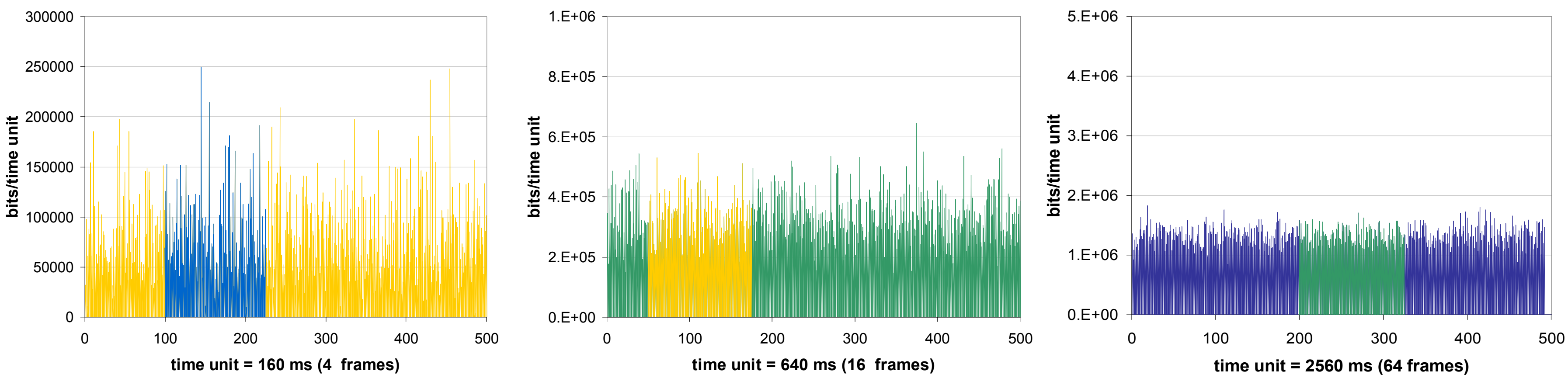
- Genuine MPEG traffic trace



W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Trans. Netw.*, vol. 2, no 1, pp. 1-15, Feb. 1994.

Self-similarity: influence of time-scales

- Synthetically generated Poisson model

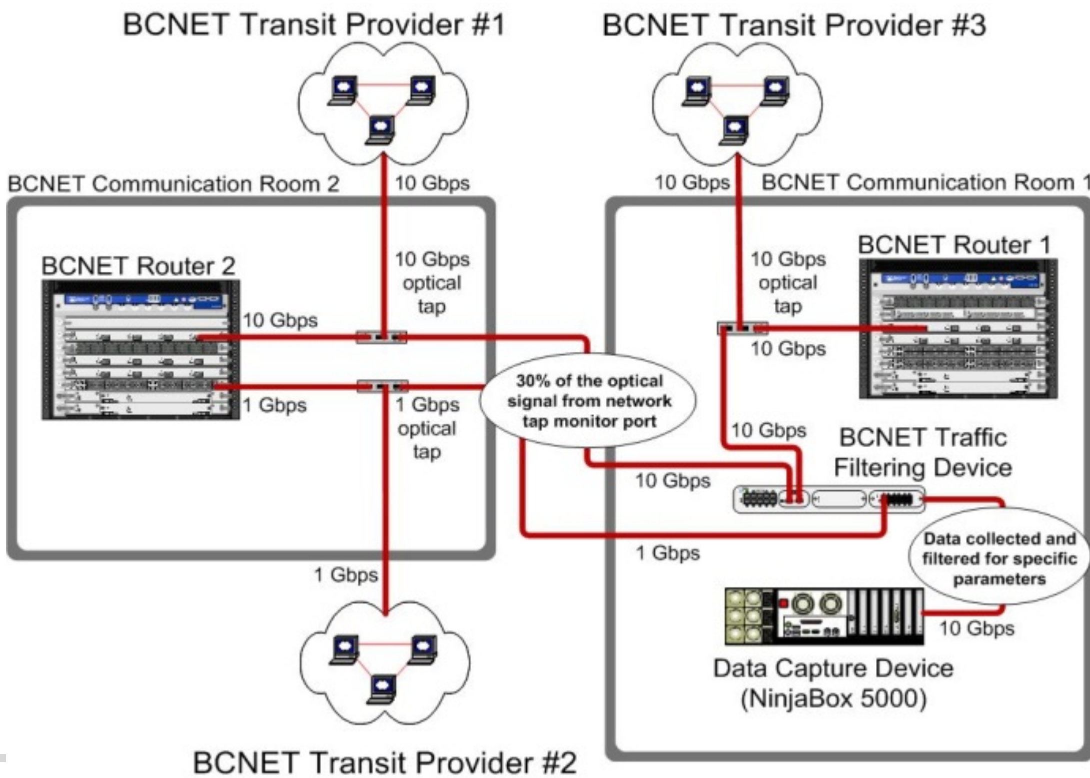


Roadmap

- Introduction
- Traffic collection, characterization, and modeling
- Case study: Collection of BCNET traffic
- Internet topology and spectral analysis of Internet graphs
- Machine learning models for feature selection and classification of traffic anomalies
- Conclusions

BCNET packet capture: physical overview

- BCNET is the hub of advanced telecommunication network in British Columbia, Canada that offers services to research and higher education institutions

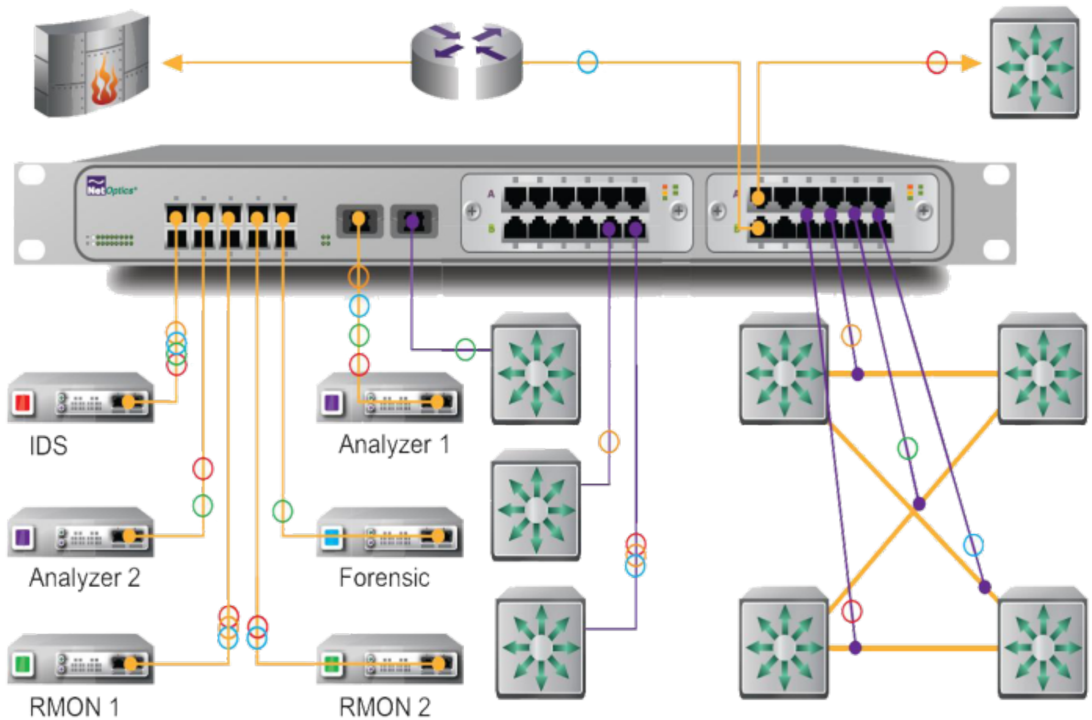


BCNET packet capture

- BCNET transits have two service providers with 10 Gbps network links and one service provider with 1 Gbps network link
- Optical Test Access Point (TAP) splits the signal into two distinct paths
- The signal splitting ratio from TAP may be modified
- The Data Capture Device (NinjaBox 5000) collects the real-time data (packets) from the traffic filtering device

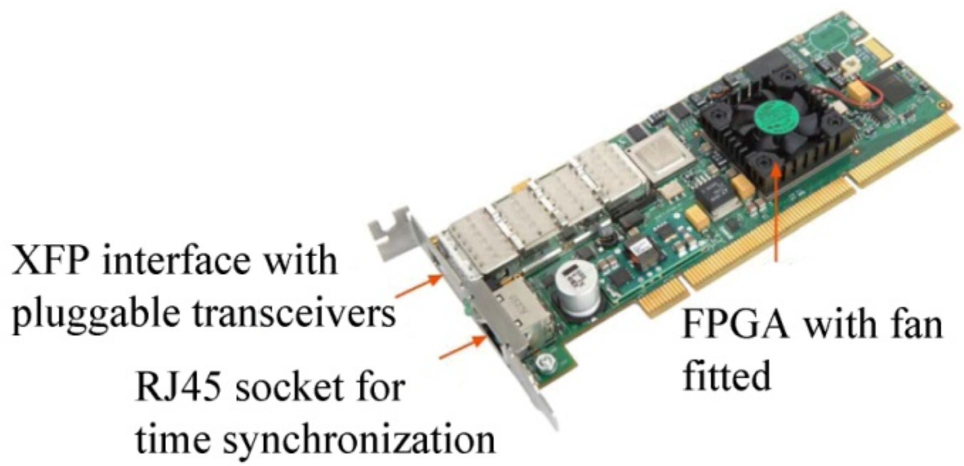
Net Optics Director 7400: application diagram

- Net Optics Director 7400 is used for BCNET traffic filtering
- It directs traffic to monitoring tools such as NinjaBox 5000 and FlowMon



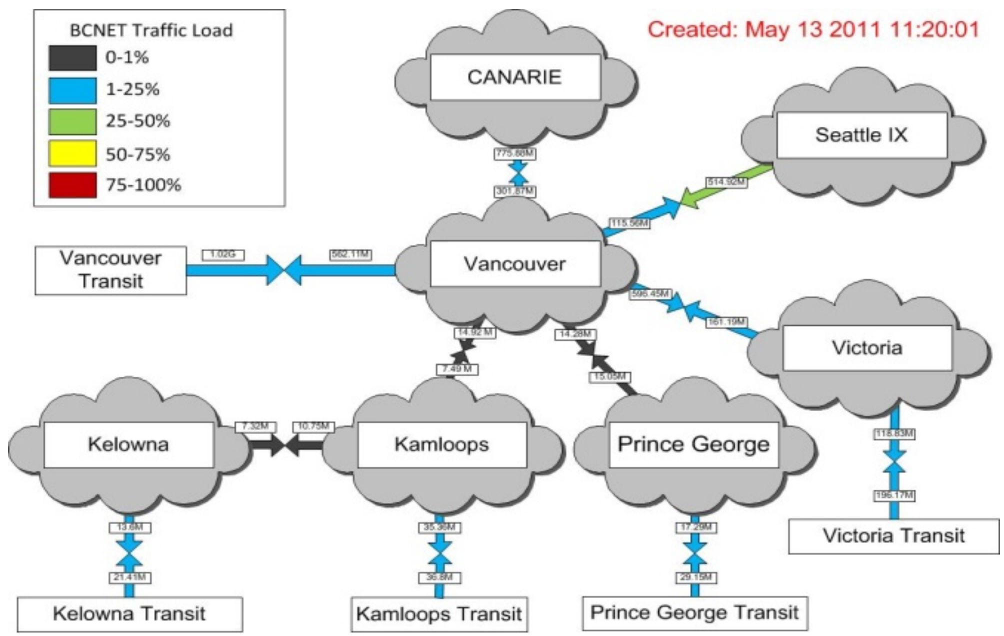
Network monitoring and analyzing: Endace card

- Endace Data Acquisition and Generation (DAG) 5.2X card resides inside the NinjaBox 5000
- It captures and transmits traffic and has time-stamping capability
- DAG 5.2X is a single port Peripheral Component Interconnect Extended (PCIx) card and is capable of capturing on average Ethernet traffic of 6.9 Gbps



Real time network usage by BCNET members

- The BCNET network is high-speed fiber optic research network
- British Columbia's network extends to 1,400 km and connects Kamloops, Kelowna, Prince George, Vancouver, and Victoria



Roadmap

- Introduction
- Traffic collection, characterization, and modeling
- Case study: Collection of BCNET traffic
- Internet topology and spectral analysis of Internet graphs
- Machine learning models for feature selection and classification of traffic anomalies
- Conclusions

Internet topology

- Internet is a network of Autonomous Systems:
  - groups of networks sharing the same routing policy
  - identified with Autonomous System Numbers (ASN)
- Autonomous System Numbers: http://www.iana.org/assignments/as-numbers
- Internet topology on AS-level:
  - the arrangement of ASes and their interconnections
- Analyzing the Internet topology and finding properties of associated graphs rely on mining data and capturing information about Autonomous Systems (ASes)

Variety of graphs

- Random graphs:
  - nodes and edges are generated by a random process
  - Erdős and Rényi model
- Small world graphs:
  - nodes and edges are generated so that most of the nodes are connected by a small number of nodes in between
  - Watts and Strogatz model (1998)

Scale-free graphs

- Scale-free graphs:
  - graphs whose node degree distribution follow power-law
  - rich get richer
  - Barabási and Albert model (1999)
- Analysis of complex networks:
  - discovery of spectral properties of graphs
  - constructing matrices describing the network connectivity

Analyzed datasets

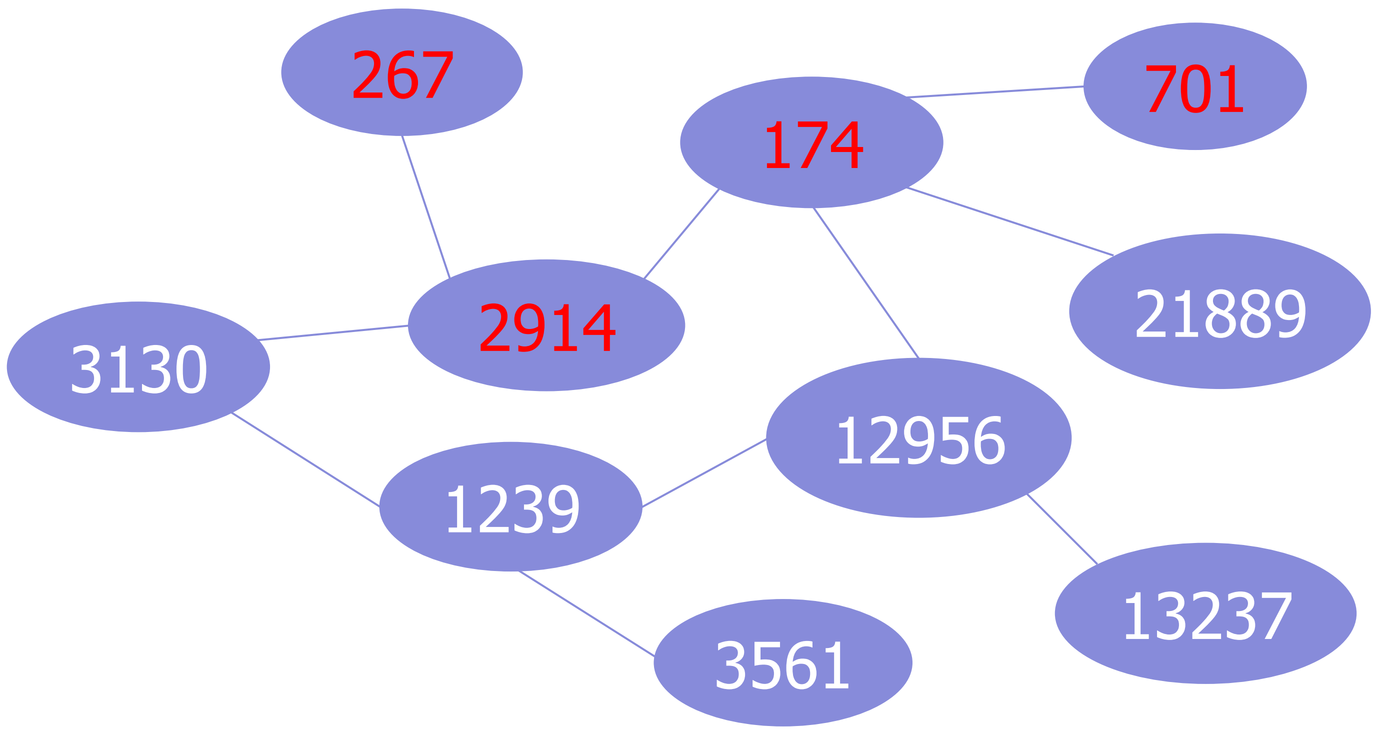
- Sample datasets:
  - Route Views:

    TABLE_DUMP| 1050122432| B| 204.42.253.253| 267| 3.0.0.0/8| 267 2914 174 701| IGP| 204.42.253.253| 0| 0| 267:2914 2914:420 2914:2000 2914:3000| NAG| |
  - RIPE:

    TABLE_DUMP| 1041811200| B| 212.20.151.234| 13129| 3.0.0.0/8| 13129 6461 7018 | IGP| 212.20.151.234| 0| 0| 6461:5997 13129:3010| NAG| |

Internet topology at AS level

- Datasets collected from Border Gateway Protocols (BGP) routing tables are used to infer the Internet topology at AS-level



Internet topology

- The Internet topology is characterized by the presence of various power-laws:
  - node degree vs. node rank
  - eigenvalues of the matrices describing Internet graphs (adjacency matrix and normalized Laplacian matrix)
- Power-laws exponents have not significantly changed over the years
- Spectral analysis reveals new historical trends and notable changes in the connectivity and clustering of AS nodes over the years

Roadmap

- Introduction
- Traffic collection, characterization, and modeling
- Case study: Collection of BCNET traffic
- Internet topology and spectral analysis of Internet graphs
- Machine learning models for feature selection and classification of traffic anomalies
- Conclusions

Traffic anomalies

- Slammer, Nimda, and Code Red I anomalies affected performance of the Internet Border Gateway Protocol (BGP)
- BGP anomalies also include: Internet Protocol (IP) prefix hijacks, miss-configurations, and electrical failures
- Techniques for detecting BGP anomalies have recently gained visible attention and importance

Anomaly detection techniques

- Classification problem:
  - assigning an "anomaly" or "regular" label to a data point
- Accuracy of a classifier depends on:
  - extracted features
  - combination of selected features
  - underlying model

Goal:

- Detect Internet routing anomalies using the Border Gateway Protocol (BGP) update messages

BGP features

Approach:

- Define a set of 37 features based on BGP update messages
- Extract the features from available BGP update messages that are collected during the time period when the Internet experienced anomalies:
  - Slammer
  - Nimda
  - Code Red I

Feature selection

- Select the most relevant features for classification using:
  - Fisher
  - Minimum Redundancy Maximum Relevance (mRMR)
  - Odds Ratio
  - Decision Tree
  - Fuzzy Rough Sets

Feature classification

- Train classifiers for BGP anomaly detection using:
  - Support Vector Machines
  - Hidden Markov Models
  - Naive Bayes
  - Decision Tree
  - Extreme Learning Machine (ELM)

BGP: update messages

- Border Gateway Protocol (BGP) enables exchange of routing information between gateway routers using update messages
- BGP update message collections:
  - Réseaux IP Européens (RIPE) under the Routing Information Service (RIS) project
  - Route Views
  - Available in multi-threaded routing toolkit (MRT) binary format

BGP: anomalies

Anomaly	Date	Duration (h)
Slammer	January 25, 2003	16
Nimda	September 18, 2001	59
Code Red I	July 19, 2001	10

Training Data	Dataset
Slammer + Nimda	Dataset 1
Slammer + Code Red I	Dataset 2
Code Red I + Nimda	Dataset 3
Slammer	Dataset 4
Nimda	Dataset 5
Code Red I	Dataset 6

Slammer worm

- Sends its replica to randomly generated IP addresses
- Destination IP address gets infected if:
  - it is a Microsoft SQL server

  or
  - a personal computer with the Microsoft SQL Server Data Engine (MSDE)

Nimda worm

- Propagates through email messages, web browsers, and file systems
- Viewing the email message triggers the worm payload
- The worm modifies the content of the web document files in the infected hosts and copies itself in all local host directories

Code Red I worm

- Takes advantage of vulnerability in the Microsoft Internet Information Services (IIS) indexing software
- It triggers a buffer overflow in the infected hosts by writing to the buffers without checking their limit

BGP: features

- Define 37 features
- Sample every minute during a five-day period:
  - the peak day of an anomaly
  - two days prior and two days after the peak day
- 7,200 samples for each anomalous event:
  - 5,760 regular samples (non-anomalous)
  - 1,440 anomalous samples
  - Imbalanced dataset

BGP features

Feature	Definition	Category
1	Number of announcements	Volume
2	Number of withdrawals	Volume
3	Number of announced NLRI prefixes	Volume
4	Number of withdrawn 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
9	Number of duplicate withdrawals	Volume
10	Number of implicit withdrawals	Volume

BGP features

Feature	Definition	Category
11	Average edit distance	AS-path
12	Maximum edit distance	AS-path
13	Inter-arrival time	Volume
14–24	Maximum edit distance = n, where n = (7, ..., 17)	AS-path
25–33	Maximum AS-path length = n, where n = (7, ..., 15)	AS-path
34	Number of IGP packets	Volume
35	Number of EGP packets	Volume
36	Number of incomplete packets	Volume
37	Packet size (B)	Volume

Feature selection algorithms

- May be employed to select the most relevant features:
  - Fisher
  - Minimum Redundancy Maximum Relevance (mRMR)
  - Odds Ratio
  - Decision Tree
  - Fuzzy Rough Sets

Feature selection: decision tree

- Commonly used algorithm in data mining
- Generates a model that predicts the value of a target variable based on several input variables
- A top-down approach is commonly used for constructing decision trees:
  - an appropriate variable is chosen to best split the set of items based on homogeneity of the target variable within subsets
- C5 software tool was used to generate decision trees

C5 [Online]. Available: http://www.rulequest.com/see5-info.html.

Feature selection: decision tree

Dataset	Training data	Selected Features
Dataset 1	Slammer + Nimda	1-21, 23-29, 34-37
Dataset 2	Slammer + Code Red I	1-22, 24-29, 34-37
Dataset 3	Code Red I + Nimda	1-29, 34-37

- Either four (30, 31, 32, 33) or five (22, 30, 31, 32, 33) features are removed in the constructed trees mainly because:
  - features are numerical and some are used repeatedly

Feature selection: fuzzy rough sets

- Deal with the approximation of fuzzy sets in a fuzzy approximation space defined by a fuzzy similarity relation or by a fuzzy partition
- The fuzzy similarity relation *Sim(C)* is:
  - an *nxn* matrix that describes similarities between any two samples
  - computed by the min operator
- Computational complexity: $O(n^2m)$
  - *n* is the number of samples
  - *m* is the number of features

Feature selection: fuzzy rough sets

Dataset	Training data	Selected Features
Dataset 4	Slammer	1, 3-6, 9, 10, 13-32, 35
Dataset 5	Nimda	1, 3-4, 8-10, 12, 14-32, 35, 36
Dataset 6	Code Red I	3-4, 8-10, 12, 14-32, 35, 36

- Using combination of datasets, for example Slammer + Nimda for training leads to higher computational load
- Each dataset was used individually

Anomaly classifiers: decision tree

Dataset	Testing data	Acc_{train}	Acc_{test}	Training time (s)
Dataset 1	Code Red I	90.7	78.8	1.8
Dataset 2	Nimda	92.3	72.8	2.1
Dataset 3	Slammer	87.1	23.8	2.3

- Each path from the root node to a leaf node may be transformed into a decision rule
- A set of rules that are obtained from a trained decision tree may be used for classifying unseen samples

Anomaly classifier: ELM

- Used for learning with a single hidden layer feed forward neural network
- Weights connecting the input and hidden layers with the bias terms are initialized randomly
- Weights connecting the hidden and output layers are analytically determined
- Learns faster than SVMs by a factor of thousands
- Suitable for online applications
- We use all features (37), all continuous features (17), features selected by fuzzy rough sets (28 or 27), and continuous features selected by fuzzy rough sets (9 or 8)

Anomaly classifiers: ELM

No. of features	Dataset	Acc_{train}	Acc_{test}	Training time (s)
37	Dataset 1	83.57 ± 0.11	80.01 ± 0.07	2.3043
	Dataset 2	83.53 ± 0.12	79.75 ± 0.08	2.2756
	Dataset 3	80.82 ± 0.09	21.65 ± 1.93	2.2747
17	Dataset 1	84.50 ± 0.07	79.91 ± 0.01	1.9268
	Dataset 2	84.43 ± 0.12	79.53 ± 0.10	1.5928
	Dataset 3	83.06 ± 0.07	51.56 ± 16.38	1.8882

- 195 hidden units
- The binary features 14–33 are removed to form a set of 17 features

Anomaly classifiers: ELM

No. of features	Dataset	Acc_{train}	Acc_{test}
28	Dataset 4	83.08 ± 0.11	80.03 ± 0.06
28 (from 37)	Dataset 5	83.08 ± 0.09	79.78 ± 0.07
27	Dataset 6	80.05 ± 0.00	81.00 ± 1.41
9	Dataset 4	84.59 ± 0.07	80.00 ± 0.05
9 (from 17)	Dataset 5	84.25 ± 0.11	79.79 ± 0.12
8	Dataset 6	83.38 ± 0.04	49.24 ± 12.90

Roadmap

- Introduction
- Traffic collection, characterization, and modeling
- Case study: Collection of BCNET traffic
- Internet topology and spectral analysis of Internet graphs
- Machine learning models for feature selection and classification of traffic anomalies
- Conclusions

Conclusions

- Data collected from deployed networks are used to:
  - evaluate network performance
  - characterize and model traffic (inter-arrival and call holding times)
  - identify trends in the evolution of the Internet topology
  - classify traffic and network anomalies

Conclusions

- Machine learning algorithms (feature selection and classification algorithms) are used for detecting BGP anomalies
- Performance of classifiers greatly depended on the employed datasets
- Feature selection algorithms were used to improve the performance of classifiers
- For smaller datasets, performance of the ELM classifier was improved by using fuzzy rough sets
- Both decision tree and ELM are relatively fast classifiers with satisfactory accuracy

References: BGP anomaly detection

- S. Deshpande, M. Thottan, T. K. Ho, and B. Sikdar, "An online mechanism for BGP instability detection and analysis," *IEEE Trans. Computers*, vol. 58, no. 11. 1470–1484, Nov. 2009.
- T. Ahmed, B. Oreshkin, and M. Coates, "Machine learning approaches to network anomaly detection," in Proc. *USENIX Workshop on Tackling Computer Systems Problems with Machine Learning Techniques*, Cambridge, MA, USA, May 2007, pp. 1–6.
- J. Li, D. Dou, Z. Wu, S. Kim, and V. Agarwal, "An Internet routing forensics framework for discovering rules of abnormal BGP events," *SIGCOMM Comput. Commun. Rev.*, vol. 35, no. 4, pp. 55–66, Oct. 2005.

References: sources of data

- RIPE RIS raw data [Online]. Available: http://www.ripe.net/data-tools/.
- University of Oregon Route Views project [Online]. Available: http://www.routeviews.org/.
- CAIDA: Center for Applied Internet Data Analysis: {Online}. Available: http://www.caida.org/home/.

References: machine learning

- J. R. Quinlan, “Induction of decision trees,” *Machine Learning*, vol. 1, no. 1, pp. 81–106, Mar. 1986.
- Z. Pawlak, “Rough sets,” *Int. J. Inform. and Comput. Sciences*, vol. 11, no. 5, pp. 341–356, Oct. 1982.
- G. B. Huang, Q. Y. Zhu, and C. K. Siew, “Extreme learning machine: theory and applications,” *Neurocomputing*, vol. 70, pp. 489–501, Dec. 2006.
- G. B. Huang, X. J. Ding, and H. M. Zhou, “Optimization method based extreme learning machine for classification,” *Neurocomputing*, vol. 74, no. 1–3, pp. 155–163, Dec. 2010.

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

- 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," to be presented at *IEEE International Conference on Systems, Man, and Cybernetics, SMC 2014*, San Diego, CA, October 2014.
- N. Al-Rousan, S. Haeri, and Lj. Trajković, "Feature selection for classification of BGP anomalies using Bayesian models," in *Proc. 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.
- T. Farah, S. Lally, R. Gill, N. Al-Rousan, R. Paul, D. Xu, and Lj. Trajković, "Collection of BCNET BGP traffic," in *Proc. 23rd ITC*, San Francisco, CA, USA, Sept. 2011, pp. 322-323.
- S. Lally, T. Farah, R. Gill, R. Paul, N. Al-Rousan, and Lj. Trajković, "Collection and characterization of BCNET BGP traffic," in *Proc. 2011 IEEE Pacific Rim Conf. Communications, Computers and Signal Processing*, Victoria, BC, Canada, Aug. 2011, pp. 830-835.