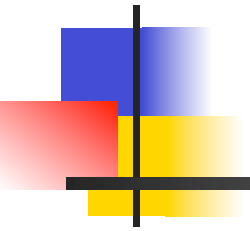
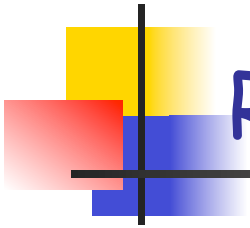


Communication Networks: Traffic Data, Network Topologies, and Routing Anomalies

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

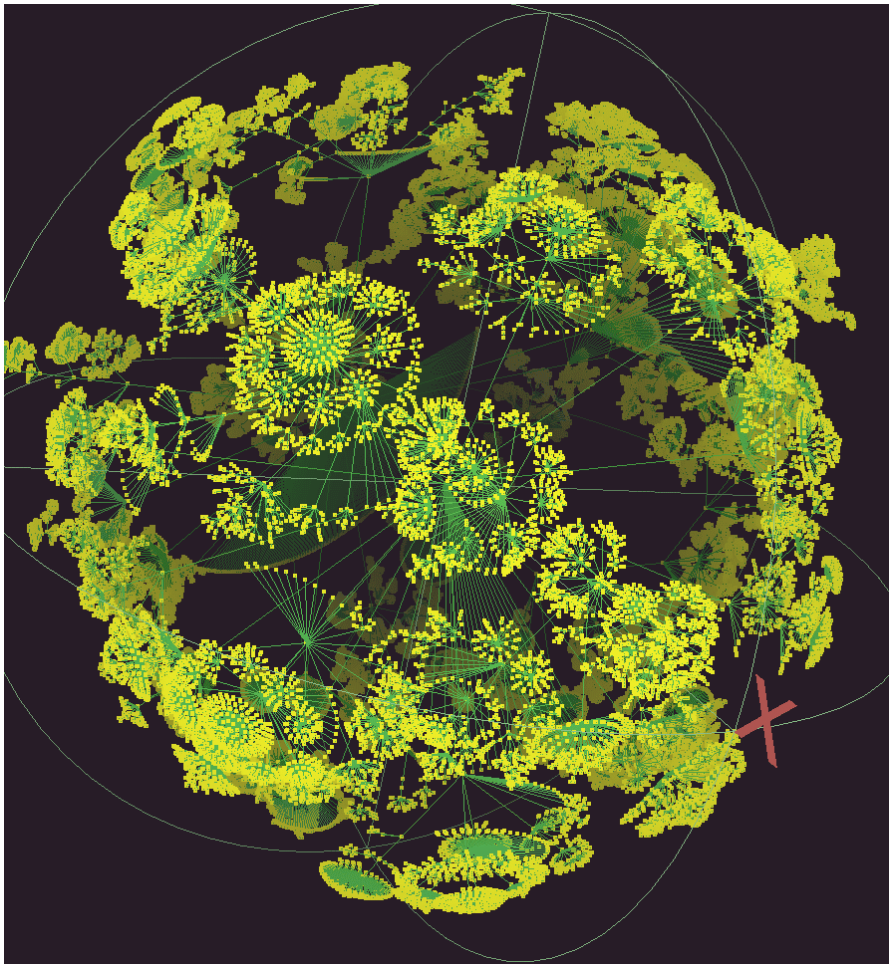
Simon Fraser University, Burnaby Campus



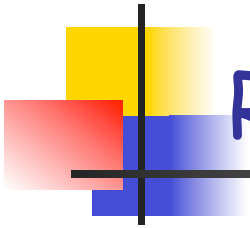
Roadmap

- 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/

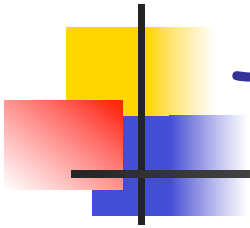


Roadmap

- 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

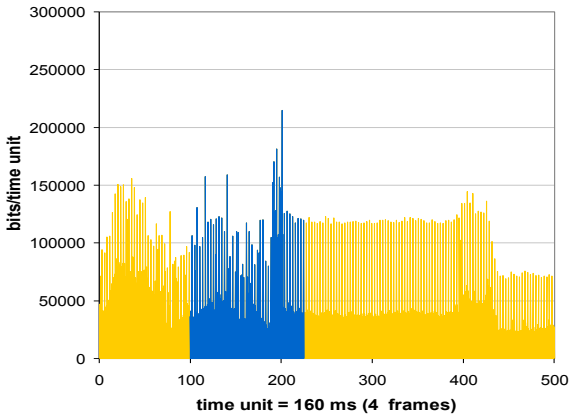


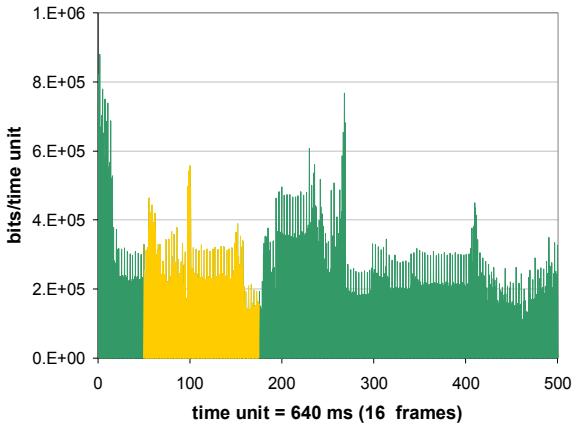
Traffic modeling: self-similarity

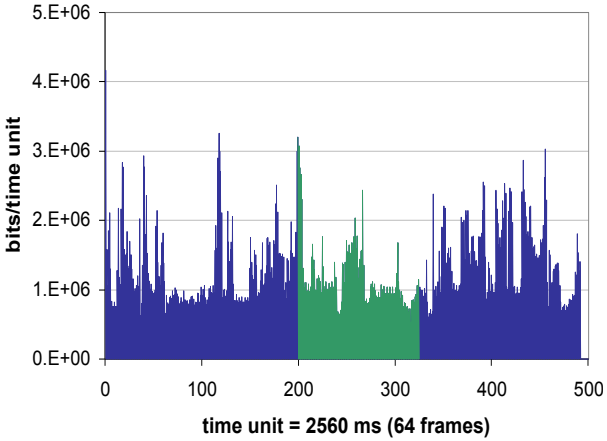
- Self-similarity implies a "fractal-like" behavior
- Data on various time scales have similar patterns
- Implications:
  - no natural length of bursts
  - bursts exist across many time scales
  - traffic does not become "smoother" when aggregated
  - it is unlike Poisson traffic used to model traffic 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



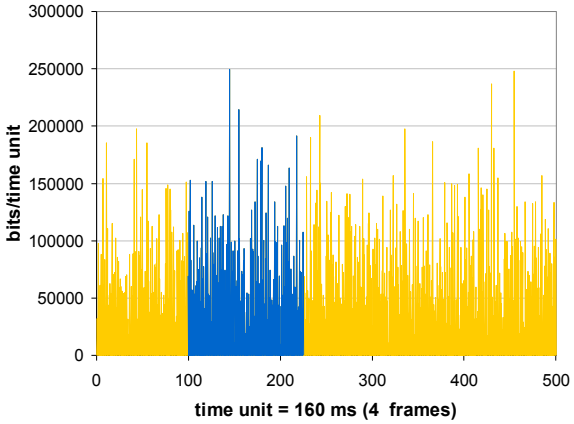


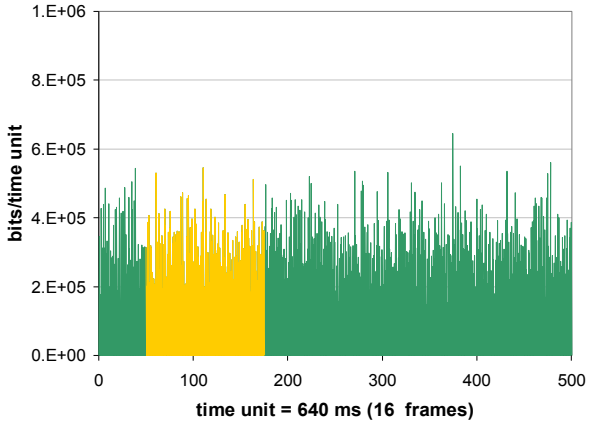


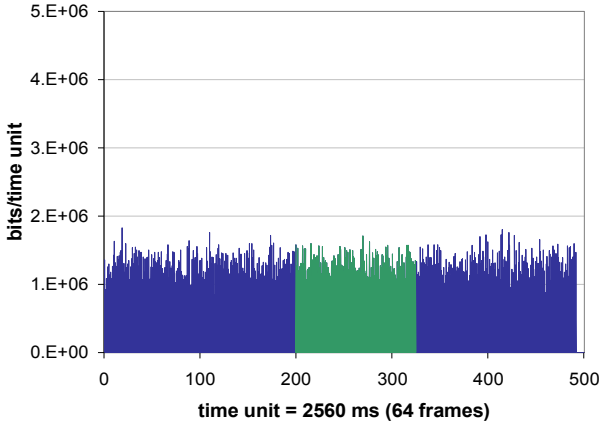
W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Trans. Netw.*, vol. 2, no 1, pp. 1-15, Feb. 1994.

Self-similarity: influence of time-scales

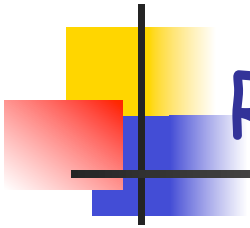
- Synthetically generated Poisson model







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.

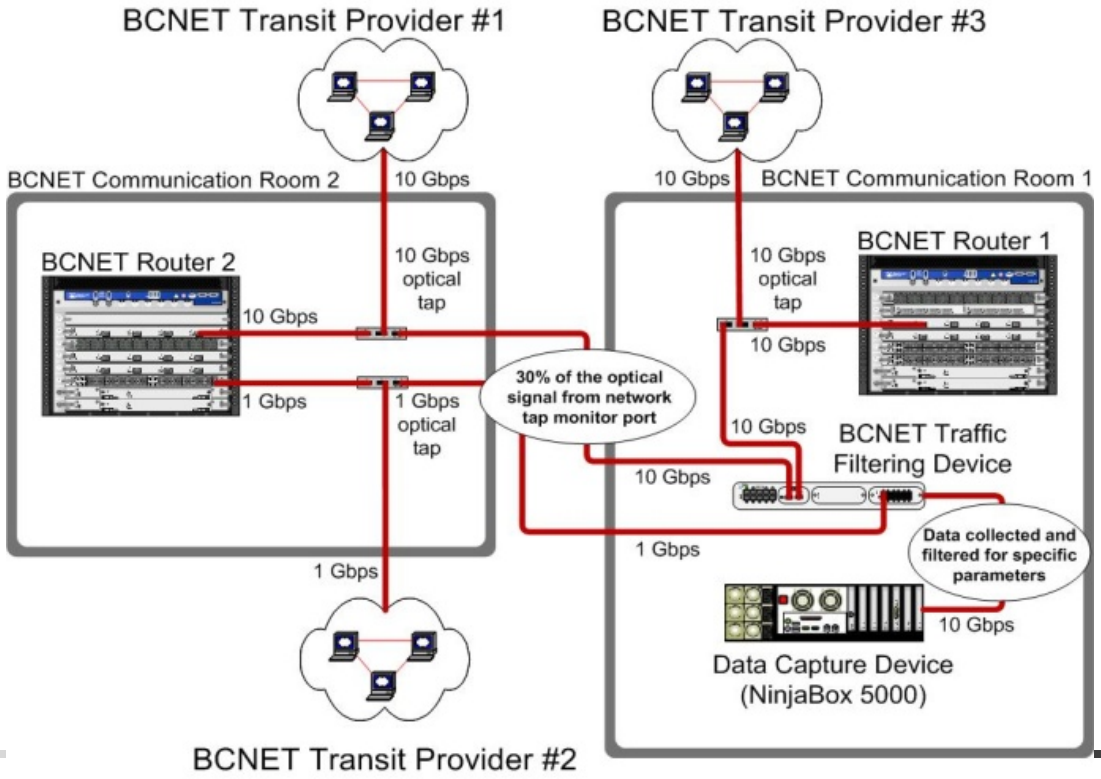


Roadmap

- 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

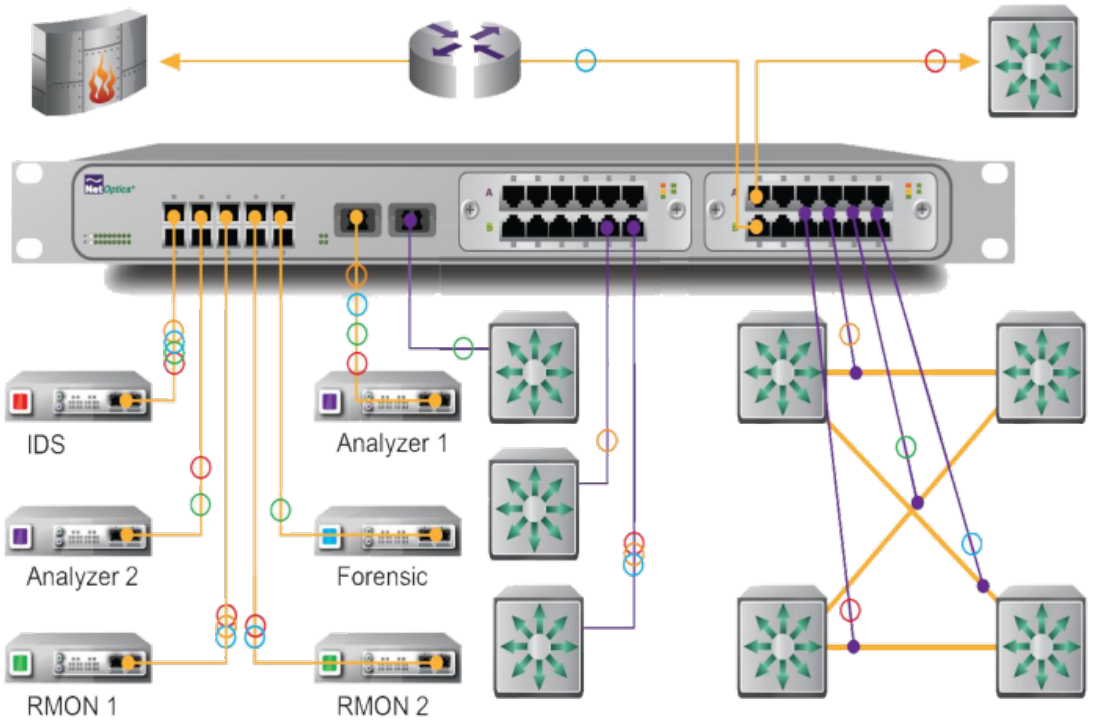


BCNET packet capture

- 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

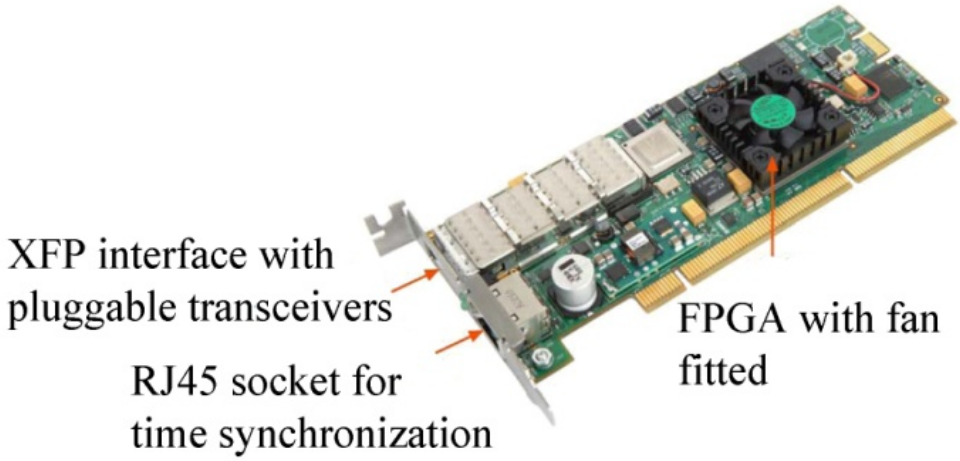
Net Optics Director 7400: application diagram

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



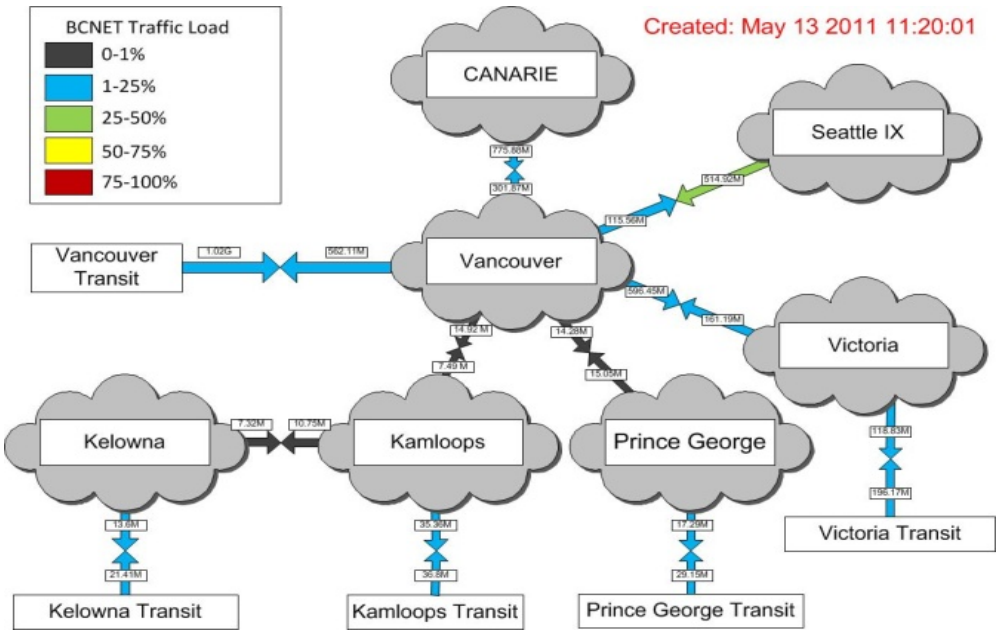
Network monitoring and analyzing: Endace card

- 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



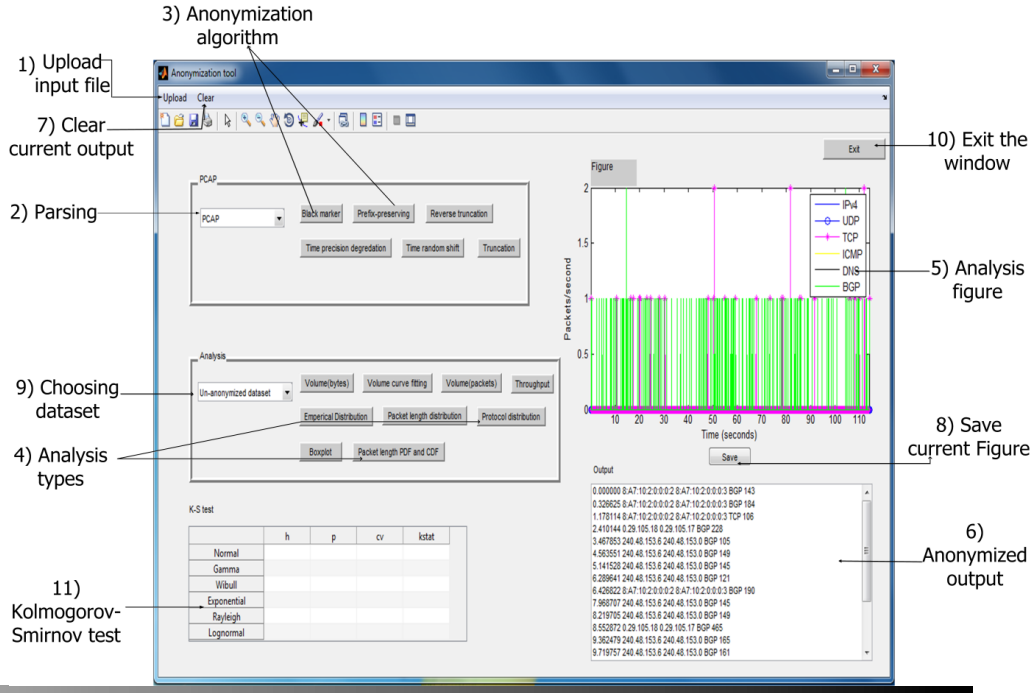
Real time network usage by BCNET members

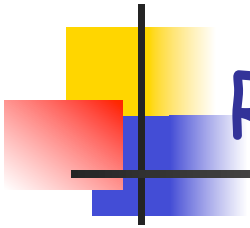
- The BCNET network is high-speed fiber optic research network
- British Columbia's network extends to 1,400 km and connects Kamloops, Kelowna, Prince George, Vancouver, and Victoria



Anonym tool

- The Anonym tool provides options to anonymize time, IPv4 and IPv6 addresses, MAC addresses, and packet length data
- Provides options to analyze the datasets

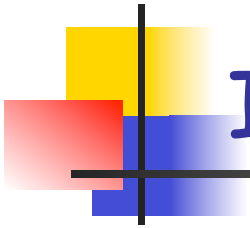




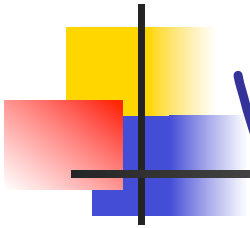
Roadmap

- 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

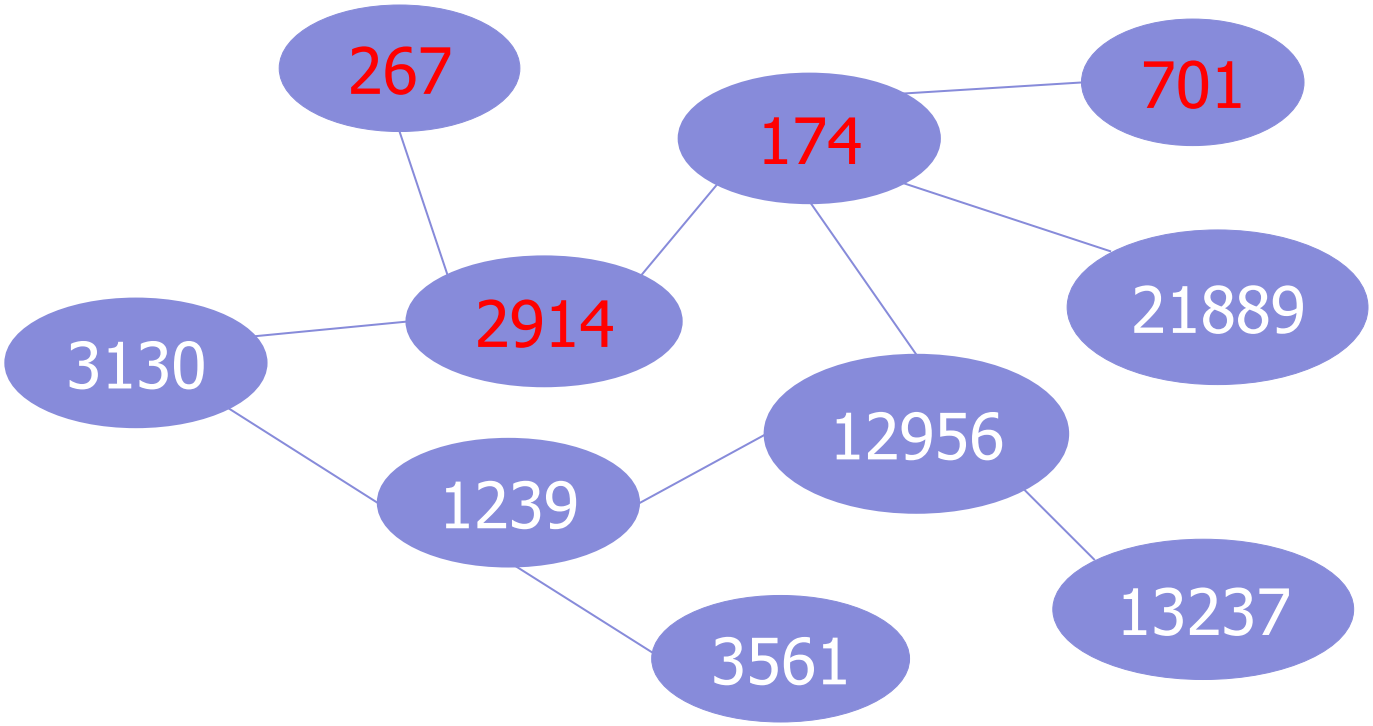
- Sample datasets:
  - Route Views:

    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| |

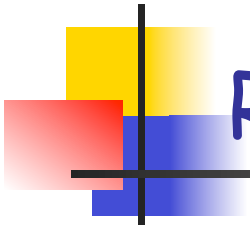
Internet topology at AS level

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



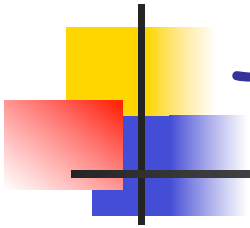
Internet topology

- 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

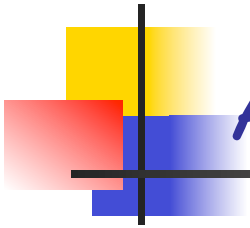


Traffic anomalies

- 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

Detection techniques

- Statistical pattern recognition:
  - main disadvantage: difficulty in estimating distributions of high dimensions
- Rule-based:
  - require a priori knowledge of network conditions
  - example: the Internet Routing Forensics (IRF)
  - they are not adaptable learning mechanisms, slow, and have
  - high degree of computational complexity

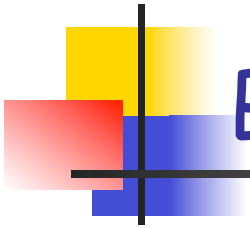


Anomaly detection techniques

- 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

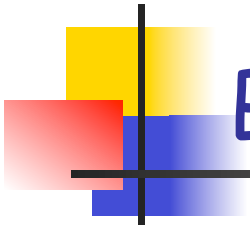
Goal:

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



BGP messages

- BGP protocol generates four types of messages: open, update, keepalive, and notification
- Only BGP update messages were considered
- BGP update messages: announcements or withdrawals



BGP features

Approach:

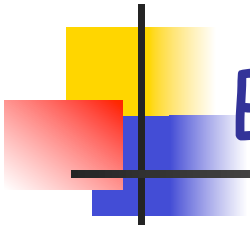
- Define a set of 37 features based on BGP update messages
- Extract the features from available BGP update messages that are collected during the time period when the Internet experienced anomalies:
  - Slammer
  - Nimda
  - Code Red I

Feature selection algorithms

- Select the most relevant features for classification using features scoring algorithms:
  - Fisher
  - Minimum Redundancy Maximum Relevance (mRMR)
  - Odds Ratio (OR) and extended/weighted/multi-class odds ratio (EOR/WOR/MOR)
  - Class discriminating measure (CDM)
  - Decision Tree
  - Fuzzy Rough Sets
- These algorithms measure the correlation and relevancy among features

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

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
- Features are normalized to have zero mean and unit variance
- This normalization reduces the effect of the Internet growth

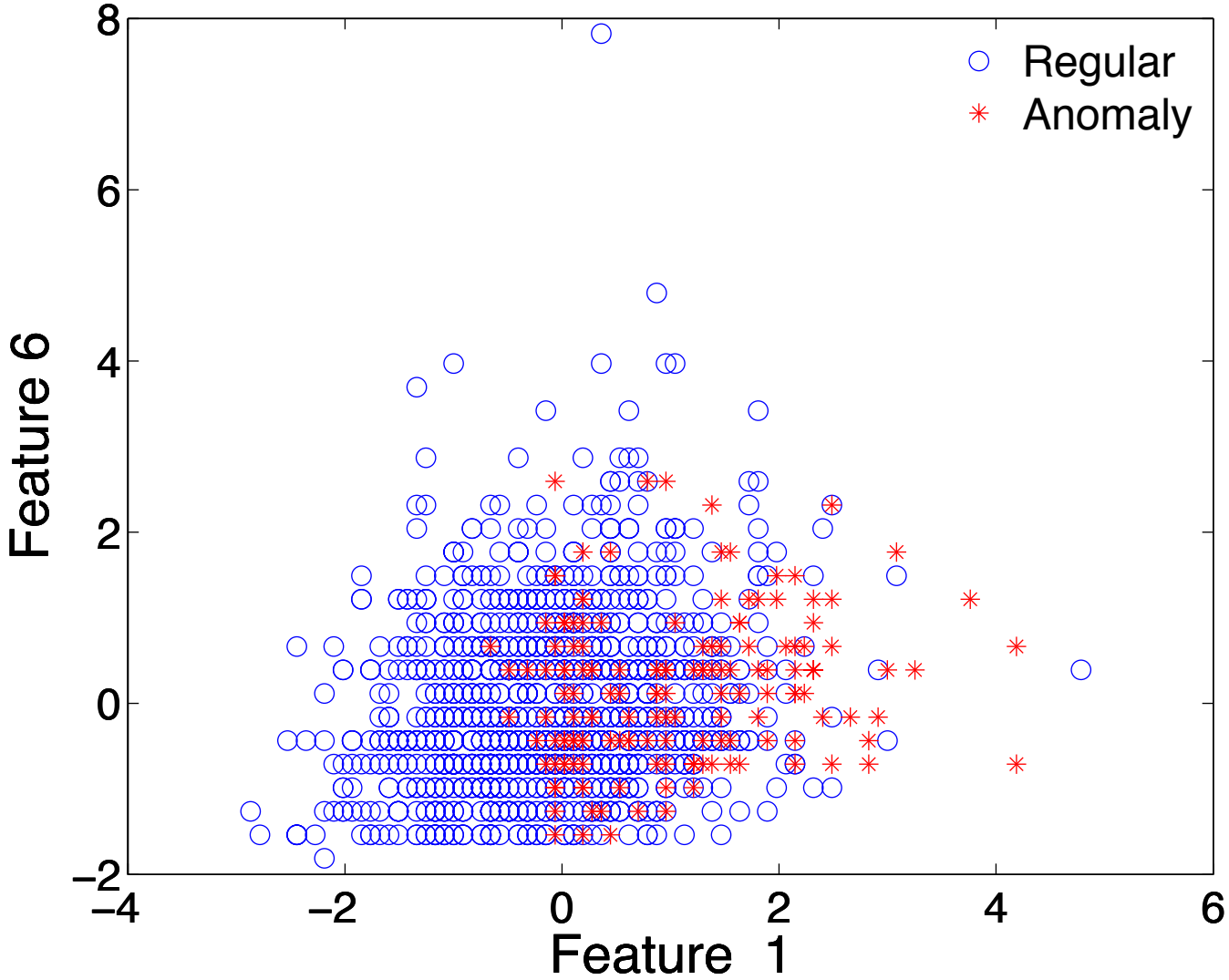
BGP features

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

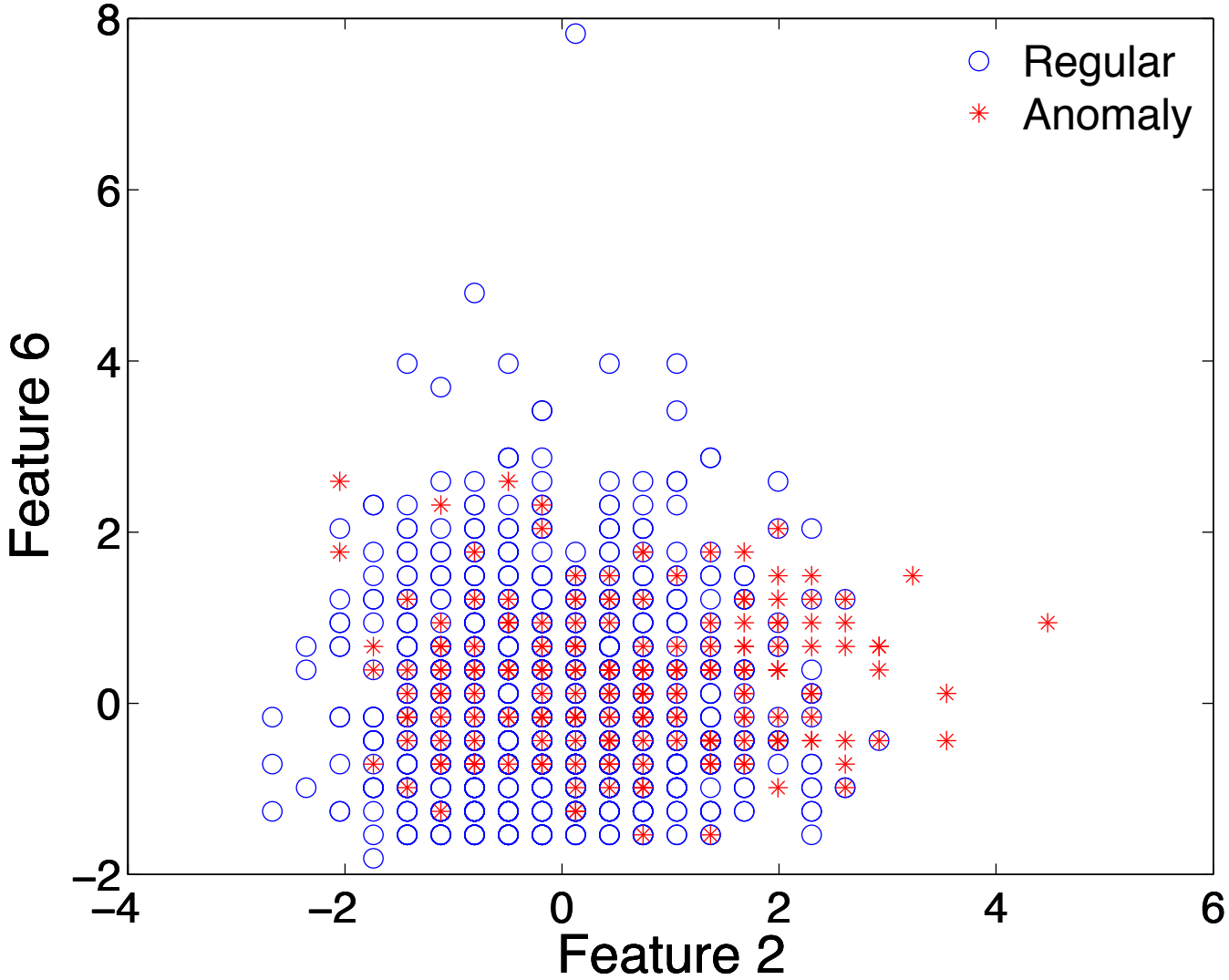
BGP features

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

Normalized scattering graphs



Normalized scattering graphs



Top ten selected features

Fisher	mRMR			Odds Ratio and its Variants				
	MID	MIQ	MIBASE	OR	EOR	WOR	MOR	CMD
11	15	15	15	10	5	5	6	5
6	5	12	17	4	7	7	5	11
9	12	3	2	1	6	6	11	9
2	7	8	8	14	11	11	17	2
16	4	1	6	12	10	13	16	8
17	10	6	3	3	4	9	14	16
8	8	4	1	15	13	2	1	3
3	13	17	9	8	9	16	2	14
1	2	9	12	17	1	17	7	1
14	14	2	11	16	14	8	3	17

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

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)

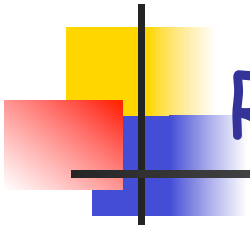
Anomaly classifiers: ELM

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

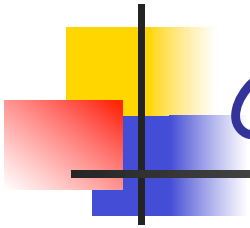
Anomaly classifiers: ELM

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
  - identify trends in the evolution of the Internet topology
  - classify traffic and network anomalies

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: 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," in *Proc. IEEE International Conference on Systems, Man, and Cybernetics, SMC 2014*, San Diego, CA, October 2014, pp. 1331-1336.
- T. Farah and Lj. Trajković, "Anonym: a tool for anonymization of the Internet traffic," in *Proc. 2013 IEEE International Conference on Cybernetics* (*CYBCONF 2013*), Lausanne, Switzerland, June 2013, pp. 261-266.
- 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.

http://ccece2016.ieee.ca/
Advancing Society through Electrical and Computer Engineering

The 29th annual IEEE Canadian Conference on Electrical and Computer Engineering (CCECE 2016) will be held in Vancouver, British Columbia, Canada from May 15 to 18, 2016. CCECE is the flagship Canadian conference for researchers, students, and professionals in the area of Electrical and Computer Engineering. It is held annually in a Canadian city to disseminate research advancements and discoveries, network and exchange ideas, strengthen existing partnerships, and foster new collaborations. CCECE 2016's general theme, *Advancing Society through Electrical and Computer Engineering*, reflects the profound impact of ECE research on our daily lives. CCECE 2016's topics include:

- Bioengineering
- Communications and networks
- Control and robotics
- Computer and software techniques
- Devices, Circuits, and Systems
- Signal theory and signal processing
- Power and energy circuits and systems

Paper submission guidelines: Submitted papers must be unpublished and should not be submitted elsewhere at the same time. Accepted papers should not exceed 6 pages in two-column IEEE Transactions style. Accepted papers longer than 4 pages will be charged $100 for each extra page. Papers should be submitted as PDF files through the paper submission system (https://edas.info/N21433). All submitted papers will be peer reviewed by at least three independent reviewers.
Note: To be published in the CCECE 2016 Conference Proceedings and to be eligible for publication in IEEE Xplore®, an author of an accepted paper is required to register for the conference at the full (member or non-member) rate and the paper must be presented by an author of that paper at the conference unless the TPC Chair grants permission for a substitute presenter arranged in advance of the event and who is qualified both to present and answer questions. For authors with multiple accepted papers, one full registration is valid for one paper and the author pays an additional $200 dollars for each additional paper. However, only one copy of the proceedings will be provided.

CCECE Founding Chair
Vijay Bhargava (University of British Columbia)
Honorary Co-Chairs
Andreas Antoniu (University of Victoria)
James Cavers (Simon Fraser University)
Hermann Dommel (University of British Columbia)
General Co-Chairs
Rodney Vaughan (Simon Fraser University)
Rabab Ward (University of British Columbia)
Technical Program Co-Chairs
Shahriar Mirabbasi (University of British Columbia)
Ljiljana Trajkovic (Simon Fraser University)
Tutorials and Panels Co-Chairs
Ivan Bajic (Simon Fraser University)
Thomas Johnson (University of British Columbia Okanagan
Publication Co-Chairs
Hamed Shah-Mansouri (University of British Columbia)
Vincent Wong (University of British Columbia)
Publicity Chair
Dave Michelson (University of British Columbia)

Patronage/Exhibition Chair
Bob Gill (British Columbia Institute of Technology)
Local Arrangements Chair
Jeff Bloemink (British Columbia Institute of Technology)
Registration Chair
Cathie Lowell (CL Consulting)
Finance Co-Chairs
Steven McClain (IEEE Vancouver Section)
Lee Vishloff (Tech-Knows Services)
Treasurer
Ashfaq (Kash) Husain (Dillon Consulting)
Website and Social Media Chair:
Stephen Makonin (Simon Fraser University)

Important Dates
Tutorial proposals: December 13, 2015
Invited sessions proposals: December 27, 2015
Regular paper submission: December 27, 2015
Acceptance notifications: February 7, 2016
Final paper submission: March 6, 2016

CCECE 2016

Canadian Conference on Electrical and Computer Engineering

http://ccece2016.ieee.ca/

Important Dates

- Tutorial proposals: Dec. 13, 2015
- Invited sessions proposals: Dec. 27, 2015
- Regular paper submission: Dec. 27, 2015
- Acceptance notifications: Feb. 7, 2015
- Final paper submission: March 6, 2015