Machine Learning for Complex Networks

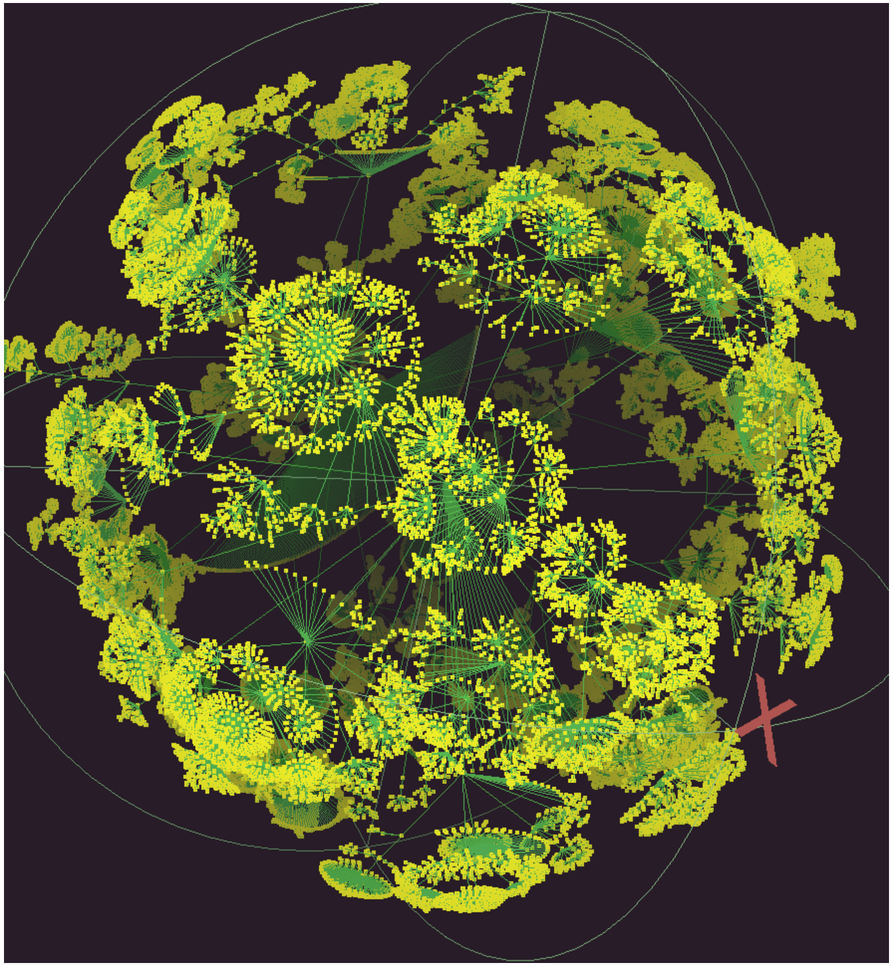
Ljiljana Trajković
ljilja@cs.sfu.ca

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 studies:
  - telecommunication network: BCNET
  - public safety wireless network: E-Comm
  - satellite network: ChinaSat
  - packet data networks: Internet
- Conclusions

Ihr: 535,102 nodes and 601,678 links



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

Roadmap

- Introduction
- Traffic collection, characterization, and modeling
- Case studies:
  - telecommunication network: BCNET
  - public safety wireless network: E-Comm
  - satellite network: ChinaSat
  - packet data networks: Internet
- 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 (unlike Poisson traffic)
  - 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

- Self-similarity implies a “fractal-like” behavior: data on various time scales have similar patterns
- A wide-sense stationary process X(n) is called (exactly second order) self-similar if its autocorrelation function satisfies:
  - $r^{(m)}(k) = r(k)$, $k \geq 0$, $m = 1, 2, \ldots, n$,

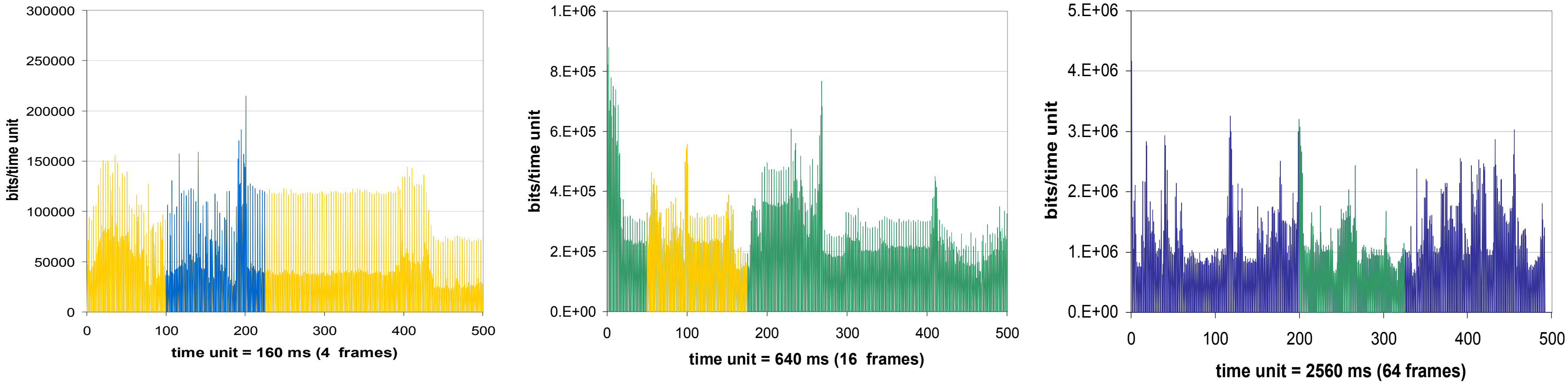
    where m is the level of aggregation

Self-similar processes

- Properties:
  - slowly decaying variance
  - long-range dependence
  - Hurst parameter (H)
- Processes with only short-range dependence (Poisson): H = 0.5
- Self-similar processes: 0.5 < H < 1.0
- As the traffic volume increases, the traffic becomes more bursty, more self-similar, and the Hurst parameter increases

Self-similarity: influence of time-scales

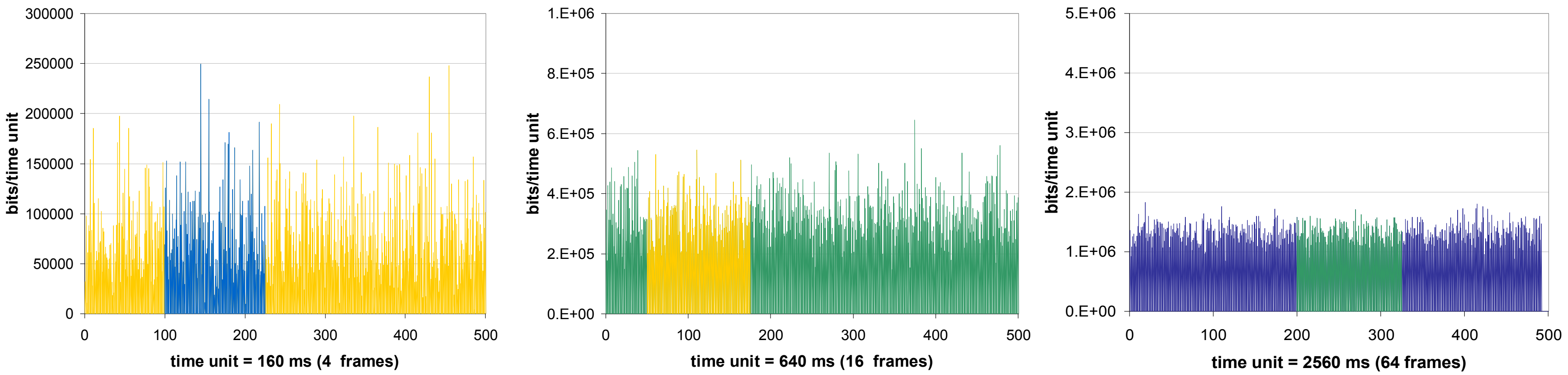
- Genuine MPEG traffic trace



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.

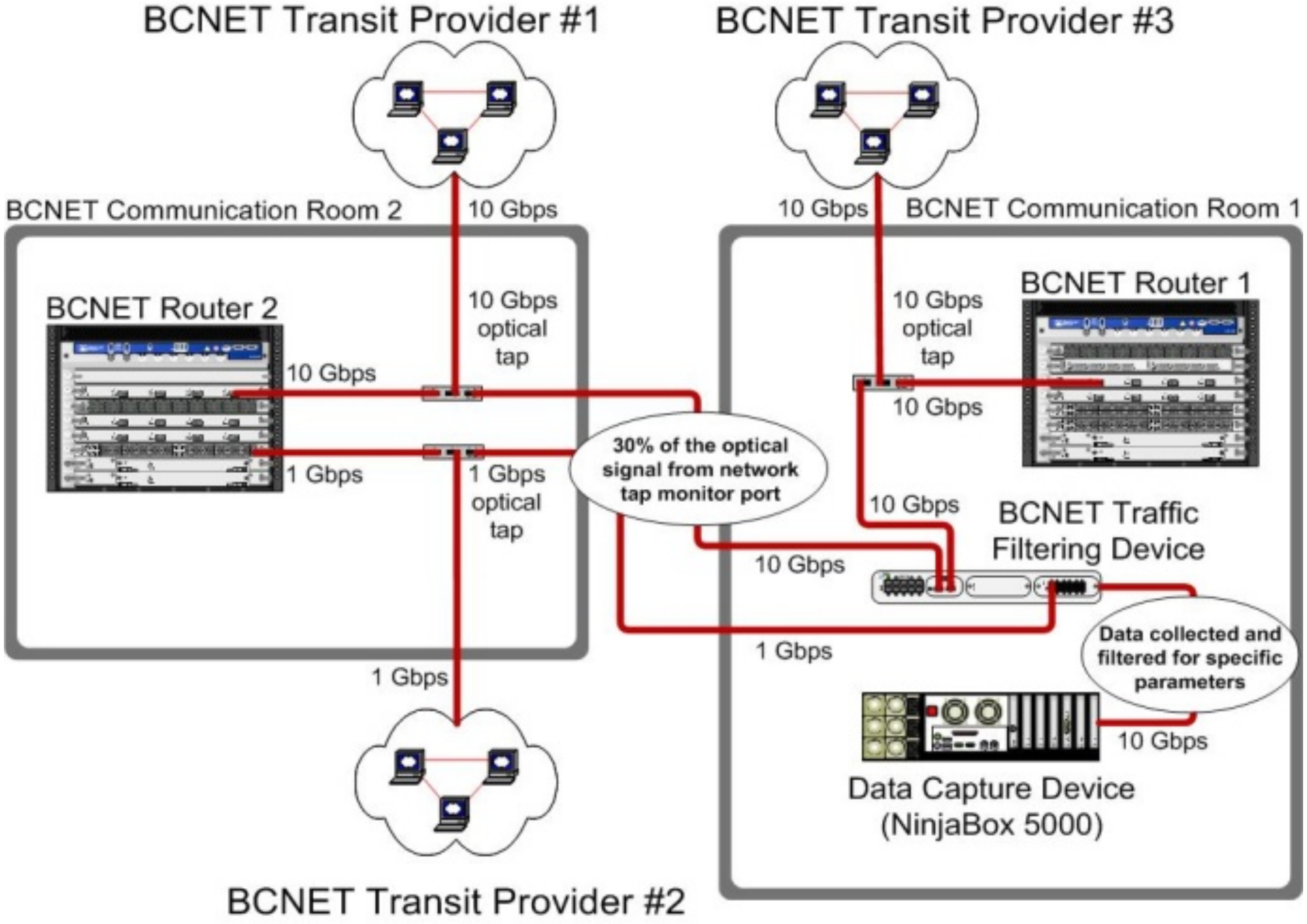
Roadmap

- Introduction
- Traffic collection, characterization, and modeling
- Case studies:
  - telecommunication network: BCNET
  - public safety wireless network: E-Comm
  - satellite network: ChinaSat
  - packet data networks: Internet
- Conclusions

Case study: BCNET

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

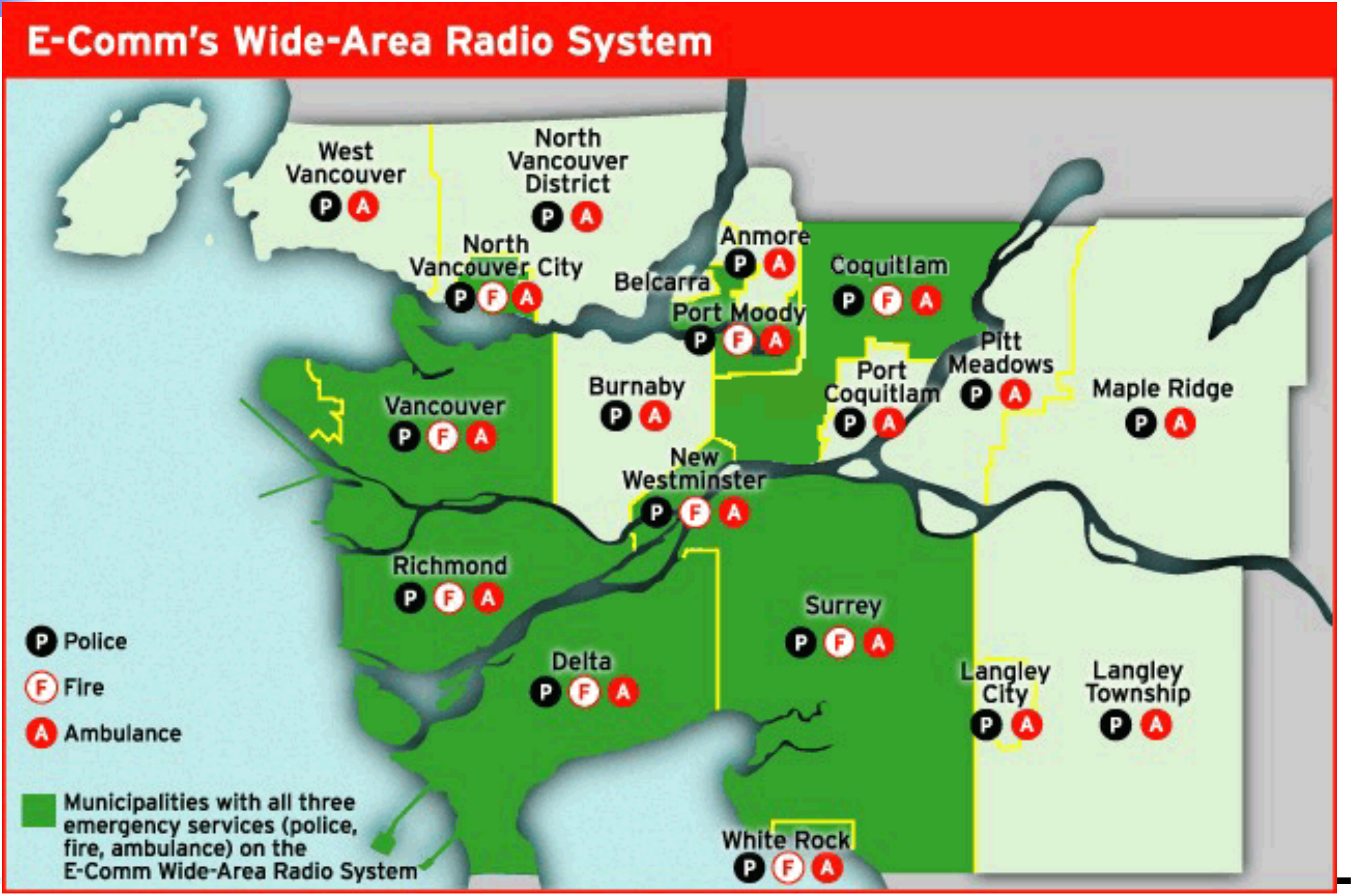
BCNET packet capture



Case study: E-Comm network

- E-Comm network: an operational trunked radio system serving as a regional emergency communication system
- The E-Comm network is capable of both voice and data transmissions
- Voice traffic accounts for over 99% of network traffic
- A group call is a standard call made in a trunked radio system
- More than 85% of calls are group calls
- A distributed event log database records every event occurring in the network: call establishment, channel assignment, call drop, and emergency call

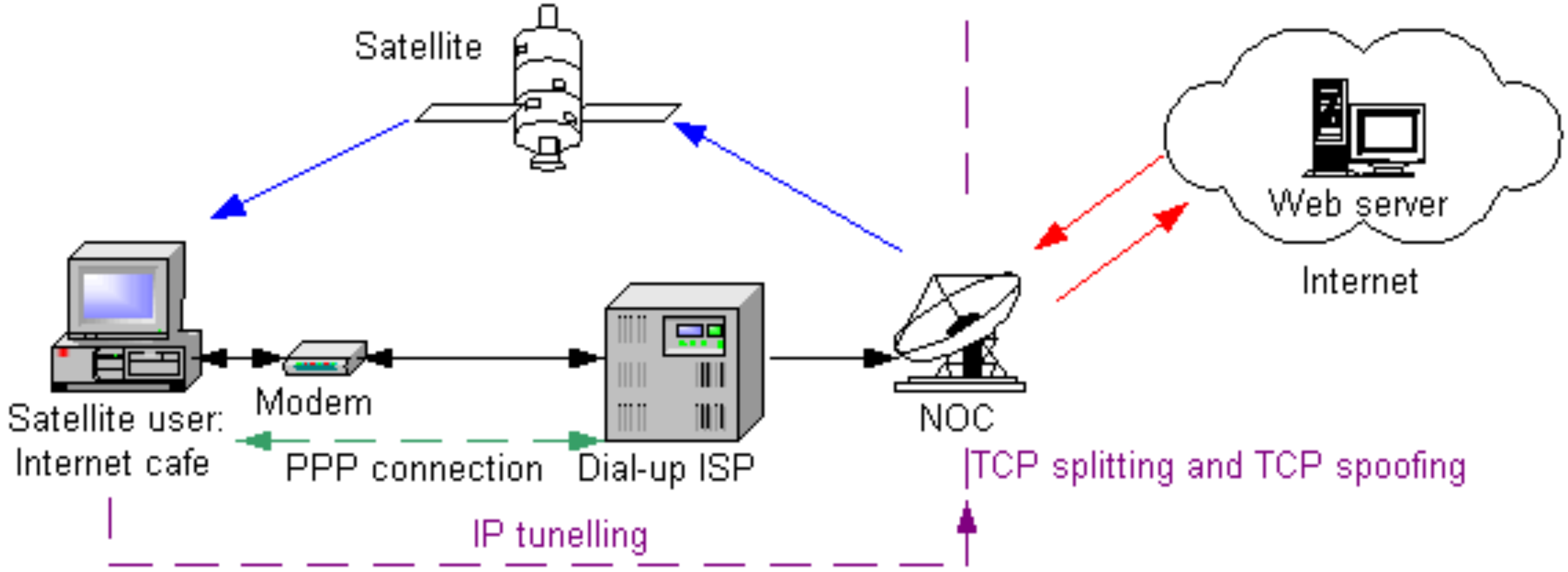
E-Comm network



Case study: ChinaSat DirecPC system

- ChinaSat hybrid satellite network
  - Employs geosynchrous satellites deployed by Hughes Network Systems Inc.
  - Provides data and television services:
    - DirecPC (Classic): unidirectional satellite data service
    - DirecTV: satellite television service
    - DirecWay (Hughnet): new bi-directional satellite data service that replaces DirecPC
  - DirecPC transmission rates:
    - 400 kb/s from satellite to user
    - 33.6 kb/s from user to network operations center (NOC) using dial-up
  - Improves performance using TCP splitting with spoofing

ChinaSat DirecPC system



Roadmap

- Introduction
- Traffic collection, characterization, and modeling
- Case studies:
  - telecommunication network: BCNET
  - public safety wireless network: E-Comm
  - satellite network: ChinaSat
  - packet data networks: Internet
- 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

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

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

Anomaly classification

- Train classifiers for BGP anomaly detection using:
  - Support Vector Machines (SVM)
  - Long Short-Term Memory (LSTM) Neural Network
  - Hidden Markov Models (HMM)
  - Naive Bayes (NB)
  - Decision Tree
  - Extreme Learning Machine (ELM)

Feature extraction: BGP 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: known anomalies

Anomaly	Date	Duration (min)
Slammer	January 25, 2003	869
Nimda	September 18-20, 2001	3,521
Code Red I	July 19, 2001	600

Event	Date	Peers
Moscow power blackout	May 2005	AS 1853, AS 12793, AS 13237
AS 9121 routing table leak	Dec. 2004	AS 1853, AS 12793, AS 13237
AS 3561 improper filtering	Apr. 2001	AS 3257, AS 3333, AS 286
Panix domain hijack	Jan. 2006	AS 12956, AS 6762, AS 6939, AS 3549
As-path error	Oct. 2001	AS 3257, AS 3333, AS 6762, AS 9057
AS 3356/AS 714 de-peering	Oct. 2005	AS 13237, AS 8342, AS 5511, AS 16034

Training and test datasets

Dataset	Training dataset	Test dataset
1	Slammer and Nimda	Code Red I
2	Slammer and Code Red I	Nimda
3	Nimda and Code Red I	Slammer
4	Slammer	Nimda and Code Red I
5	Nimda	Slammer and Code Red I
6	Code Red I	Slammer and Nimda
7	Slammer, Nimda, and Code Red I	RIPE or BCNET

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

Feature extraction: BGP messages

- 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

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

Feature selection: decision tree

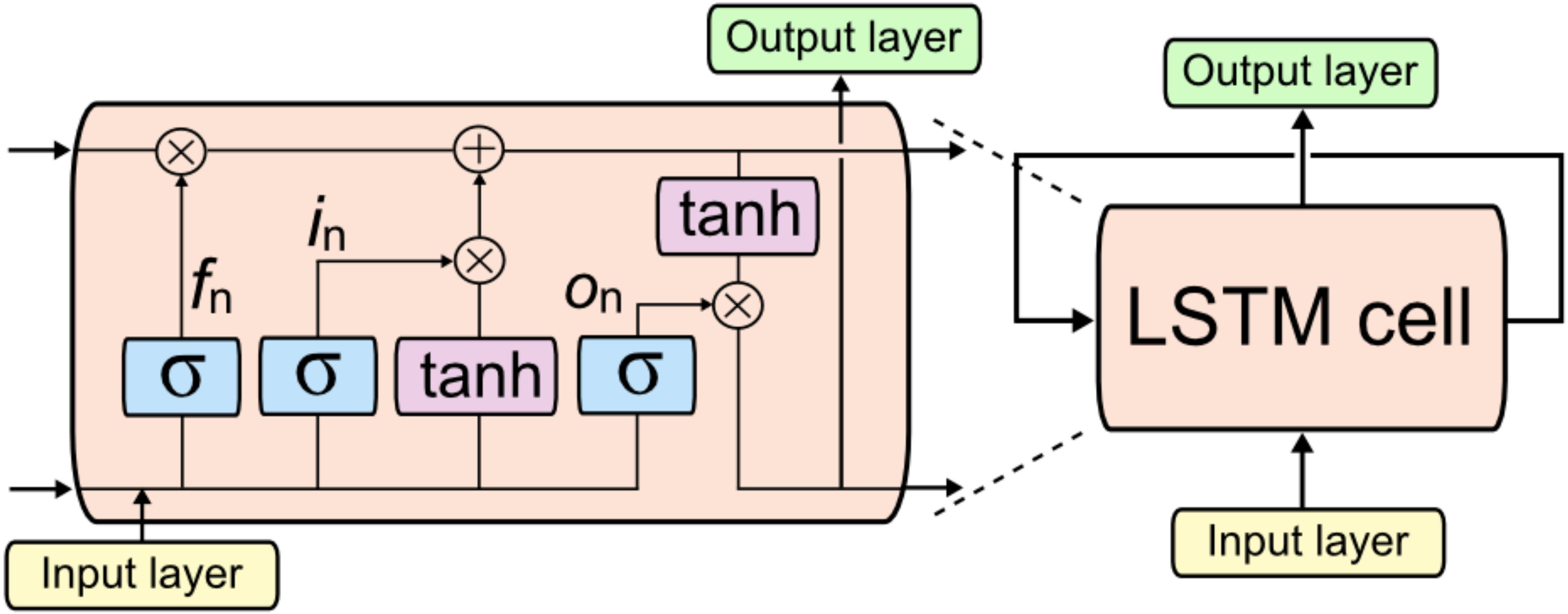
Dataset	Training data	Selected Features
Dataset 1	Slammer + Nimda	1-21, 23-29, 34-37
Dataset 2	Slammer + Code Red I	1-22, 24-29, 34-37
Dataset 3	Code Red I + Nimda	1-29, 34-37

- 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

Anomaly classification

- Train classifiers for BGP anomaly detection using:
  - Support Vector Machines (SVM)
  - Long Short-Term Memory (LSTM) Neural Network
  - Hidden Markov Models (HMM)
  - Naive Bayes (NB)
  - Decision Tree
  - Extreme Learning Machine (ELM)

Anomaly classifiers: LSTM



- Repeating modules for the LSTM neural network: input layer, LSTM layer with one LSTM cell, and output layer.

Anomaly classifiers: LSTM

		Accuracy (%)			F-Score (%)
		Test dataset	RIPE	BCNET	Test dataset
LSTMu 1	Code Red I	95.22	65.49	57.30	83.17
LSTMu 2	Nimda	53.94	51.53	50.80	11.81
LSTMu 3	Slammer	95.87	56.74	58.55	84.62

		Accuracy (%)			F-Score (%)
		Test dataset	RIPE	BCNET	Test dataset
LSTMb 1	Code Red I	56.43	60.48	62.78	26.59
LSTMb 2	Nimda	53.32	44.27	53.58	65.96
LSTMb 3	Slammer	82.98	55.00	48.20	58.54

Anomaly classifiers: decision tree

		Accuracy (%)			F-Score (%)
Training dataset	Test dataset		RIPE	BCNET	Test dataset
Dataset 1	Code Red I	85.36	89.00	77.22	47.82
Dataset 2	Nimda	58.13	94.19	81.18	26.16
Dataset 3	Slammer	95.89	89.42	77.78	84.34

- 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

Roadmap

- Introduction
- Traffic collection, characterization, and modeling
- Case studies:
  - telecommunication network: BCNET
  - public safety wireless network: E-Comm
  - satellite network: ChinaSat
  - packet data networks: Internet
- 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

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/.

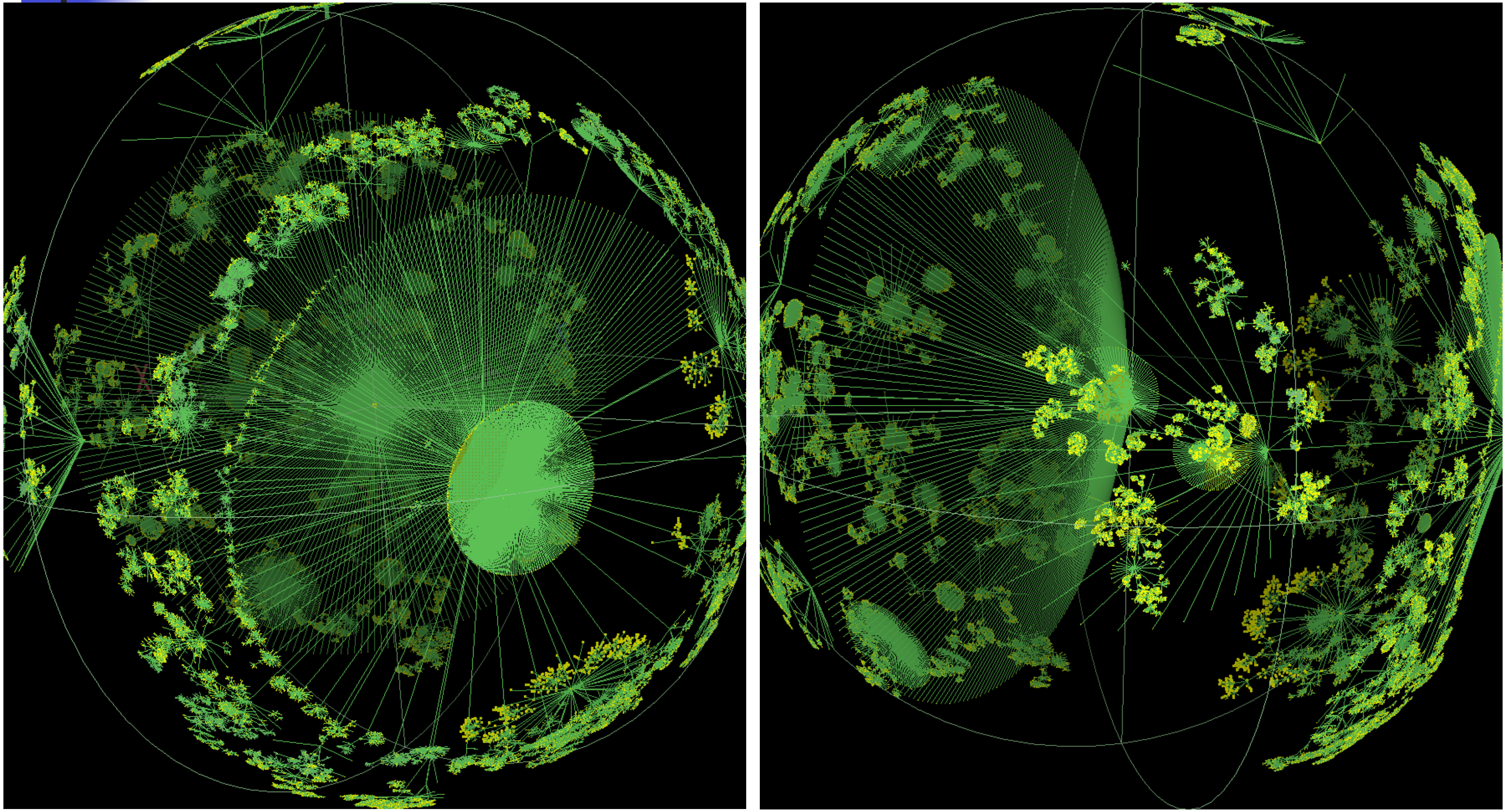
References: http://www.sfu.ca/~ljilja/cnl

- Q. Ding, Z. Li, S. Haeri, and Lj. Trajkovic, “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, to appear.
- Q. Ding, Z. Li, S. Haeri, and Lj. Trajkovic, “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, to appear.
- Q. Ding, Z. Li, P. Batta, and Lj. Trajkovic, "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.
- M. Cosovic, S. Obradovic, and Lj. Trajkovic, "Classifying anomalous events in BGP datasets," in *Proc. The 29th Annual IEEE Canadian Conference on Electrical and Computer Engineering (CCECE 2016),* Vancouver, Canada, May 2016, pp. 697-700.
- M. Cosovic, S. Obradovic, and Lj. Trajković, “Performance evaluation of BGP anomaly classifiers,” in *Proc. The Third International Conference on Digital Information, Networking, and Wireless Communications, DINWC 2015*, Moscow, Russia, Feb. 2015, pp. 115-120.

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. 1312-1317.
- T. Farah and Lj. Trajkovic, "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. 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.
- 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.

lhr: 535,102 nodes and 601,678 links



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