

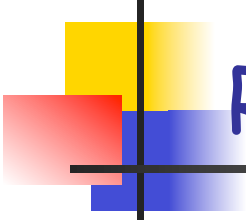
Mining Network Traffic Data

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

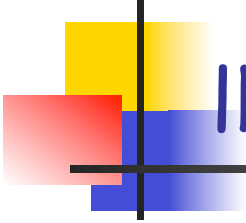
Simon Fraser University, Burnaby Campus



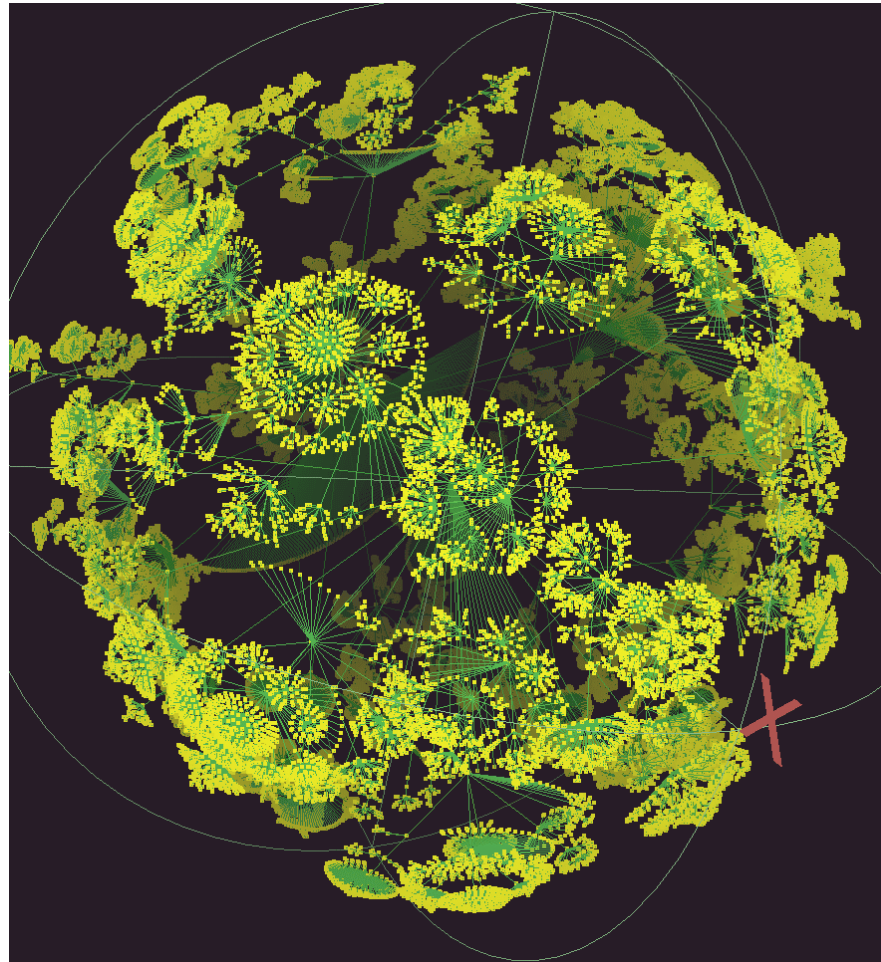


Roadmap

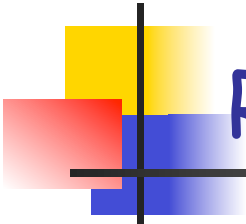
- Introduction
- Traffic collection, characterization, and modeling
- Case studies:
 - telecommunication network: BCNET
 - public safety wireless network: E-Comm
 - satellite network: ChinaSat
 - packet data networks: Internet
- Conclusions



lhr: 535,102 nodes and 601,678 links

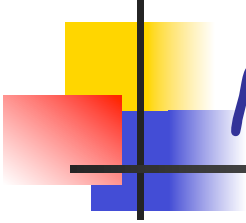


<http://www.caida.org/home/>



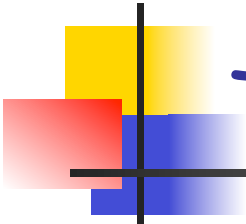
Roadmap

- 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



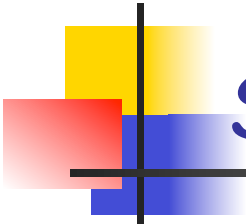
Measurements of network traffic

- **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



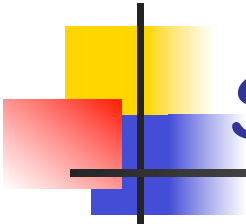
Traffic modeling: self-similarity

- 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:
 - $r^{(m)}(k) = r(k)$, $k \geq 0$, $m = 1, 2, \dots, n$,
where m is the level of aggregation

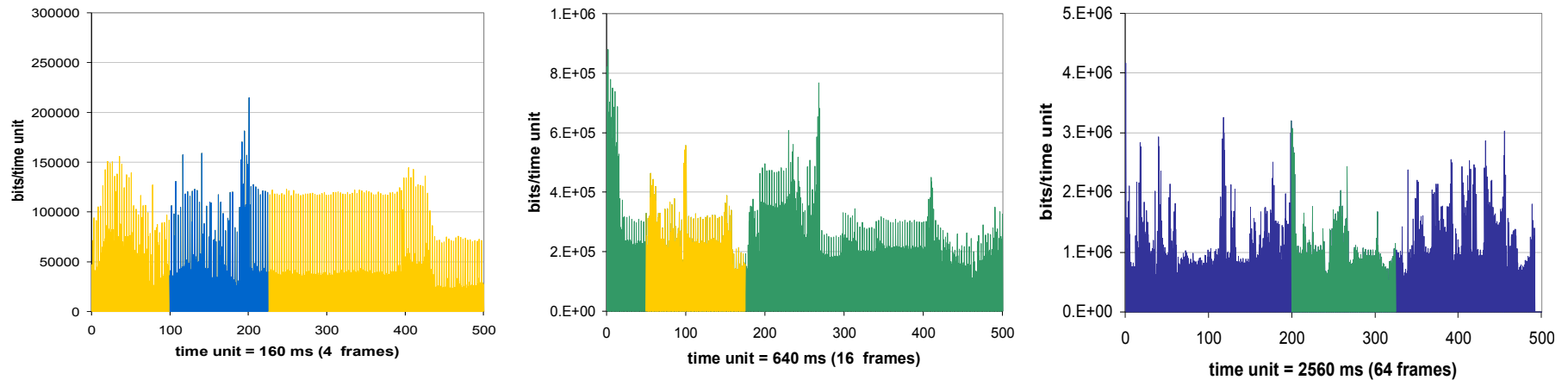


Self-similar processes

- 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

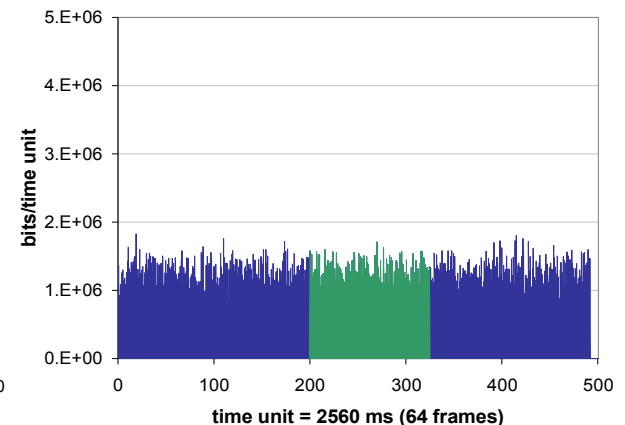
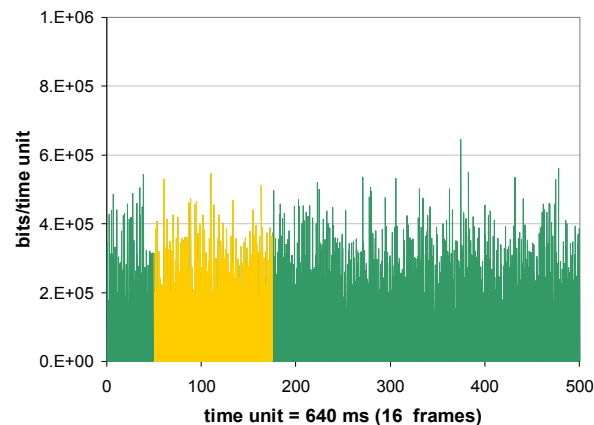
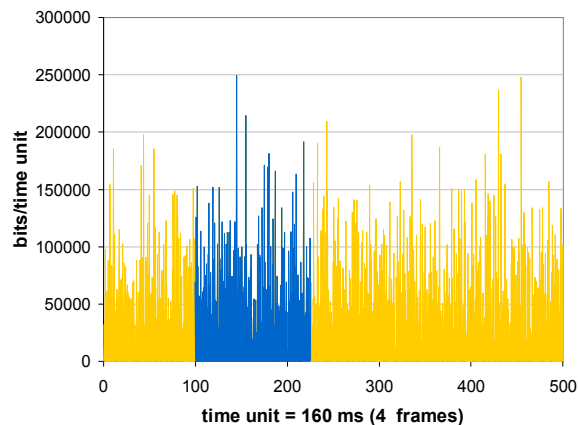
■ Genuine MPEG traffic trace



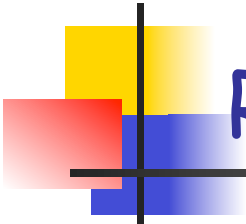
W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Trans. Netw.*, vol. 2, no 1, pp. 1-15, Feb. 1994.

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.

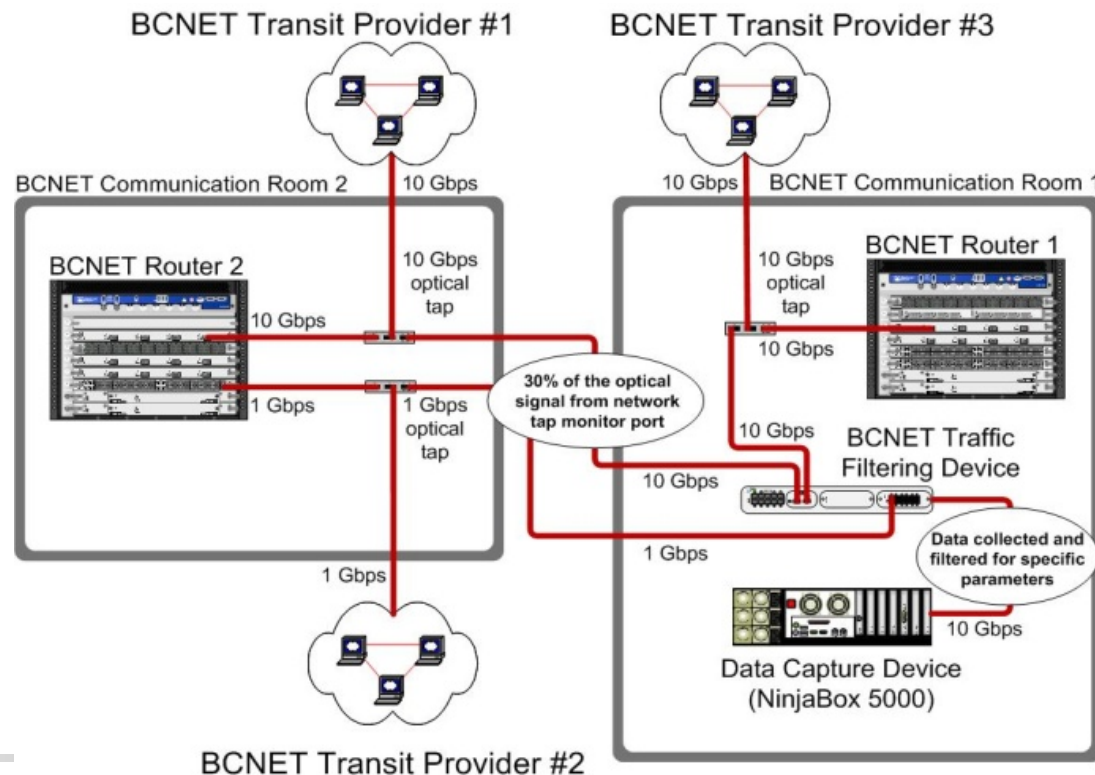


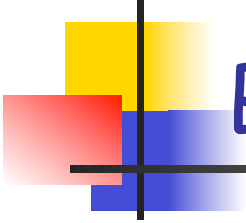
Roadmap

- 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 packet capture

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



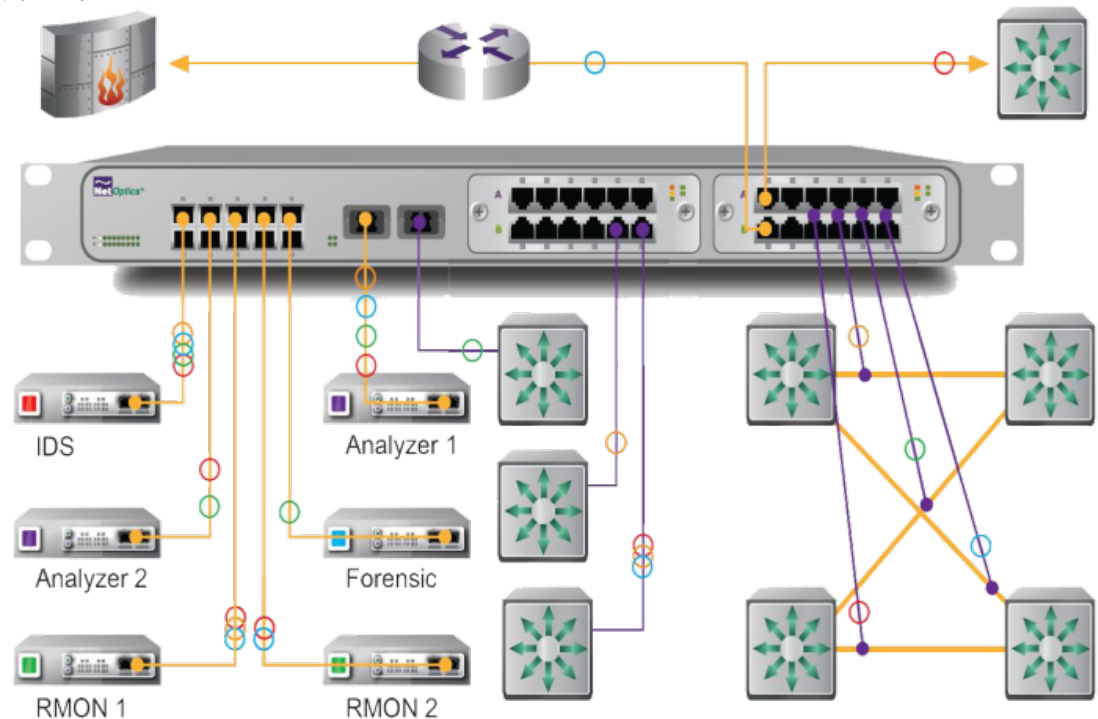


BCNET packet capture

- 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

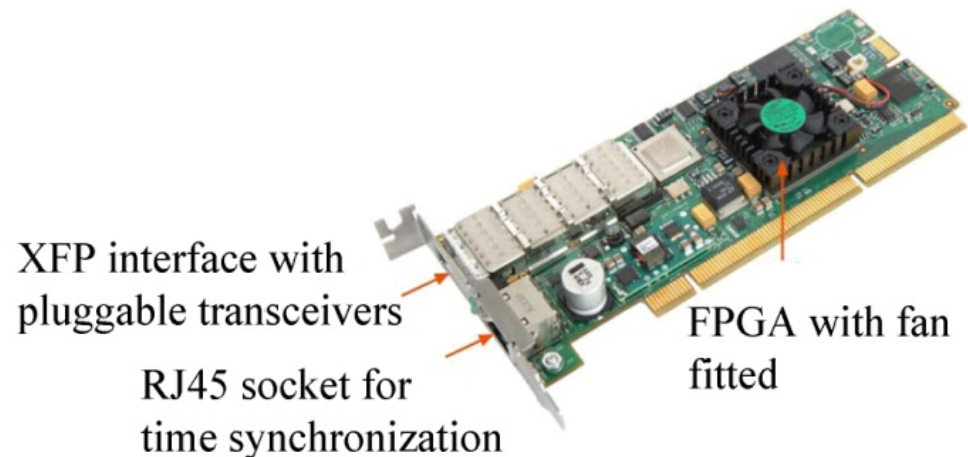
Net Optics Director 7400: application diagram

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



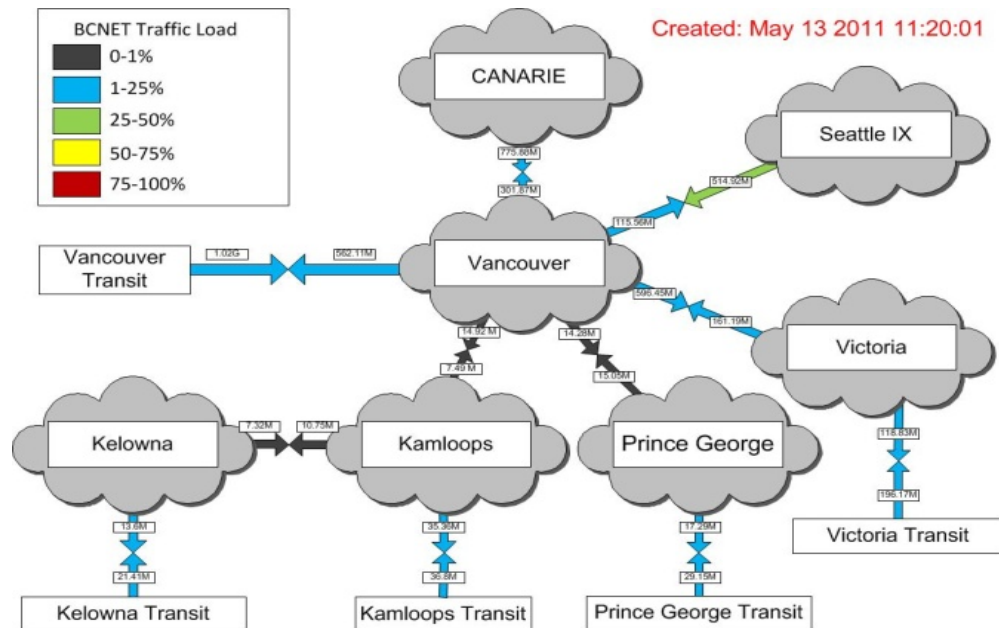
Network monitoring and analyzing: Endace card

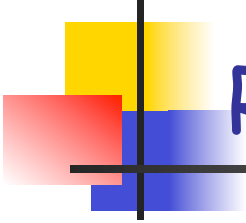
- Endace Data Acquisition and Generation (DAG) 5.2X card resides inside the NinjaBox 5000
- It captures and transmits traffic and has time-stamping capability
- DAG 5.2X is a single port Peripheral Component Interconnect Extended (PCIe) card and is capable of capturing on average Ethernet traffic of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



Real time network usage by BCNET members

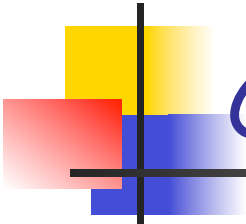
- The BCNET network is high-speed fiber optic research network
- British Columbia's network extends to 1,400 km and connects Kamloops, Kelowna, Prince George, Vancouver, and Victoria





Roadmap

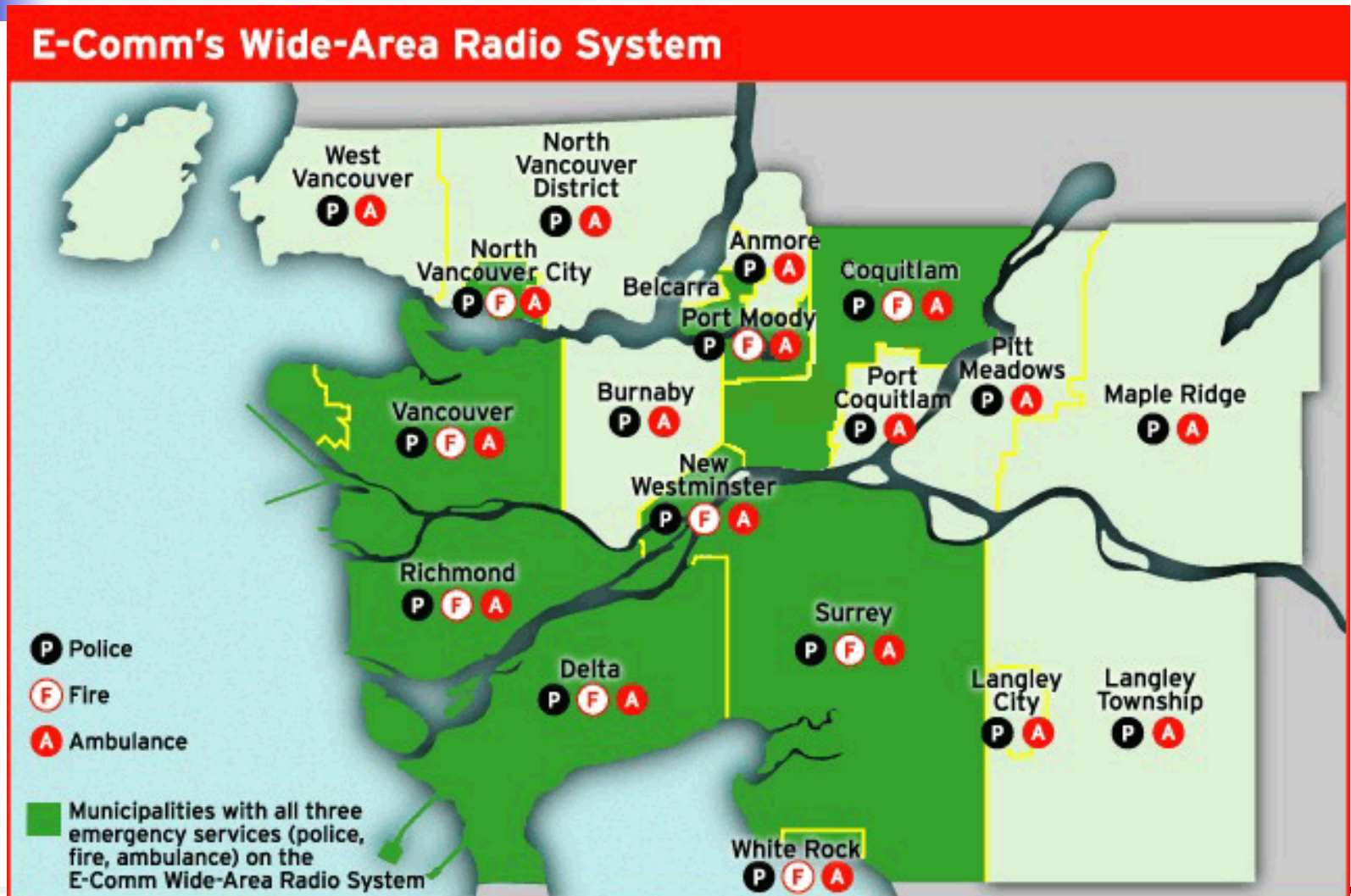
- Introduction
- Traffic collection, characterization, and modeling
- Case studies:
 - telecommunication network: BCNET
 - public safety wireless network: E-Comm
 - satellite network: ChinaSat
 - packet data networks: Internet
- Conclusions



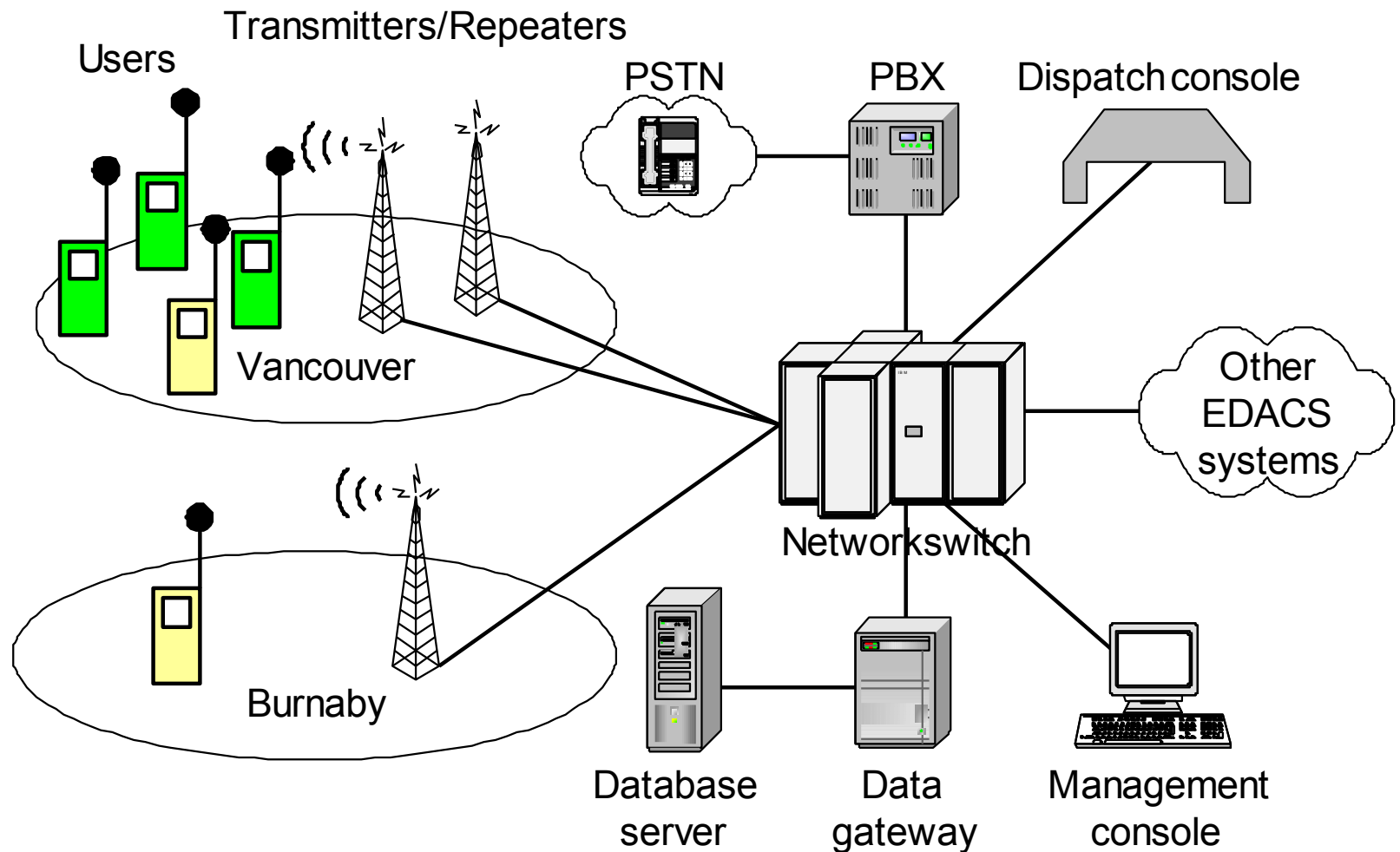
Case study: E-Comm network

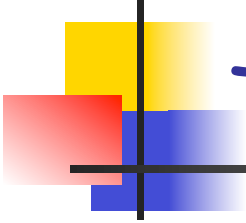
- E-Comm network: an operational trunked radio system serving as a regional emergency communication system
- The E-Comm network is capable of both voice and data transmissions
- Voice traffic accounts for over 99% of network traffic
- A group call is a standard call made in a trunked radio system
- More 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



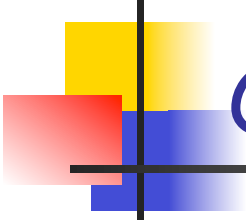
E-Comm network architecture





Traffic data

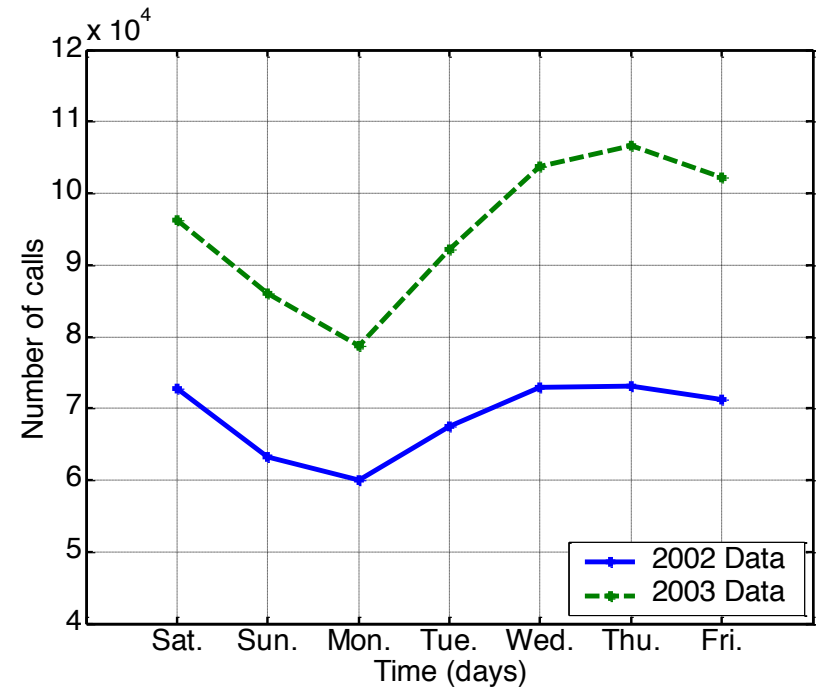
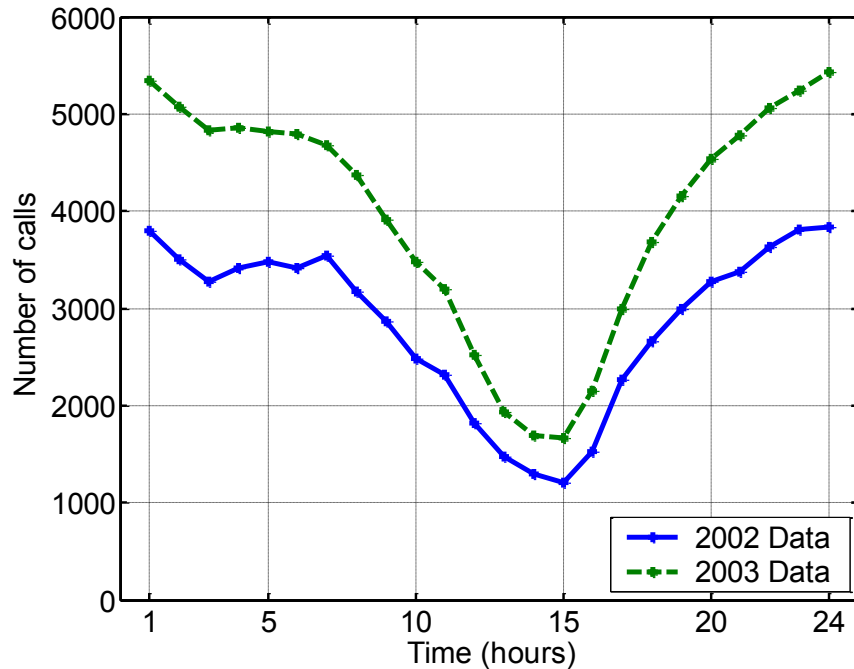
- 2001 data set:
 - 2 days of traffic data
 - 2001-11-1 to 2001-11-02 (110,348 calls)
- 2002 data set:
 - 28 days of continuous traffic data
 - 2002-02-10 to 2002-03-09 (1,916,943 calls)
- 2003 data set:
 - 92 days of continuous traffic data
 - 2003-03-01 to 2003-05-31 (8,756,930 calls)



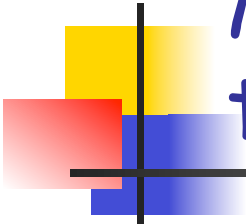
Observations

- Presence of daily cycles:
 - minimum utilization: ~ 2 PM
 - maximum utilization: 9 PM to 3 AM
- 2002 sample data:
 - cell 5 is the busiest
 - others seldom reach their capacities
- 2003 sample data:
 - several cells (2, 4, 7, and 9) have all channels occupied during busy hours

Call arrival rate in 2002 and 2003: cyclic patterns

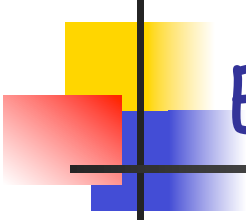


- the busiest hour is around midnight
- the busiest day is Thursday
- useful for scheduling periodical maintenance tasks



Modeling and characterization of traffic

- We analyzed **voice traffic** from a public safety wireless network in Vancouver, BC
 - call inter-arrival and call holding times during five busy hours from each year (**2001, 2002, 2003**)
- Statistical distribution and the autocorrelation function of the traffic traces:
 - Kolmogorov-Smirnov goodness-of-fit test
 - autocorrelation functions
 - wavelet-based estimation of the Hurst parameter
- B. Vujičić, N. Cackov, S. Vujičić, and Lj. Trajković, “Modeling and characterization of traffic in public safety wireless networks,” in *Proc. SPECTS 2005*, Philadelphia, PA, July 2005, pp. 214-223.



Erlang traffic models

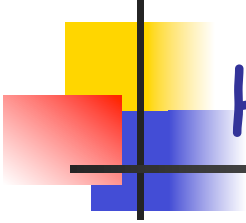
Erlang B

$$P_B = \frac{\frac{A^N}{N!}}{\sum_{x=0}^N \frac{A^x}{x!}}$$

Erlang C

$$P_C = \frac{\frac{A^N}{N!} \frac{N}{N-A}}{\sum_{x=0}^{N-1} \frac{A^x}{x!} + \frac{A^N}{N!} \frac{N}{N-A}}$$

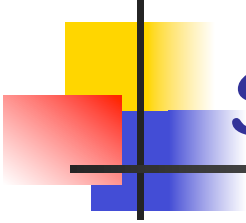
- P_B : probability of rejecting a call
- P_C : probability of delaying a call
- N : number of channels/lines
- A : total traffic volume



Hourly traces

- Call holding and call inter-arrival times from the **five busiest hours** in each dataset (2001, 2002, and 2003)

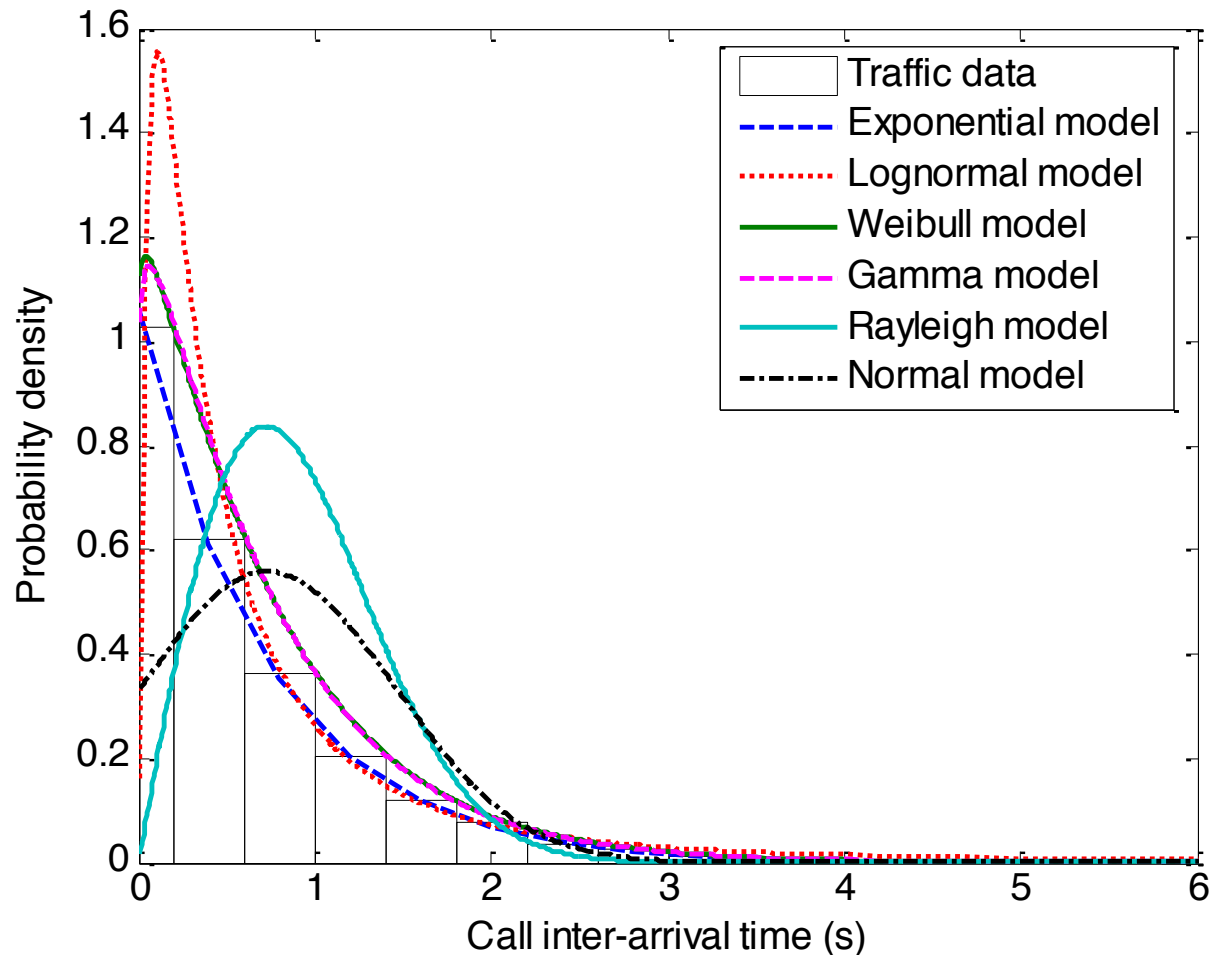
2001		2002		2003	
Day/hour	No.	Day/hour	No.	Day/hour	No.
02.11.2001 15:00–16:00	3,718	01.03.2002 04:00–05:00	4,436	26.03.2003 22:00–23:00	4,919
01.11.2001 00:00–01:00	3,707	01.03.2002 22:00–23:00	4,314	25.03.2003 23:00–24:00	4,249
02.11.2001 16:00–17:00	3,492	01.03.2002 23:00–24:00	4,179	26.03.2003 23:00–24:00	4,222
01.11.2001 19:00–20:00	3,312	01.03.2002 00:00–01:00	3,971	29.03.2003 02:00–03:00	4,150
02.11.2001 20:00–21:00	3,227	02.03.2002 00:00–01:00	3,939	29.03.2003 01:00–02:00	4,097



Statistical distributions

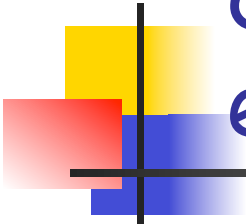
- Fourteen candidate distributions:
 - exponential, Weibull, gamma, normal, lognormal, logistic, log-logistic, Nakagami, Rayleigh, Rician, t-location scale, Birnbaum-Saunders, extreme value, inverse Gaussian
- Parameters of the distributions: calculated by performing maximum likelihood estimation
- Best fitting distributions are determined by:
 - visual inspection of the distribution of the trace and the candidate distributions
 - Kolmogorov-Smirnov test of potential candidates

Call inter-arrival times: pdf candidates



Call inter-arrival times: K-S test results (2003 data)

Distribution	Parameter	26.03.2003, 22:00–23:00	25.03.2003, 23:00–24:00	26.03.2003, 23:00–24:00	29.03.2003, 02:00–03:00	29.03.2003, 01:00–02:00
Exponential	h	1	1	0	1	1
	p	0.0027	0.0469	0.4049	0.0316	0.1101
	k	0.0283	0.0214	0.0137	0.0205	0.0185
Weibull	h	0	0	0	0	0
	p	0.4885	0.4662	0.2065	0.286	0.2337
	k	0.0130	0.0133	0.0164	0.014	0.0159
Gamma	h	0	0	0	0	0
	p	0.3956	0.3458	0.127	0.145	0.1672
	k	0.0139	0.0146	0.0181	0.0163	0.0171
Lognormal	h	1	1	1	1	1
	p	1.015E-20	4.717E-15	2.97E-16	3.267E-23	4.851E-21
	k	0.0689	0.0629	0.0657	0.0795	0.0761

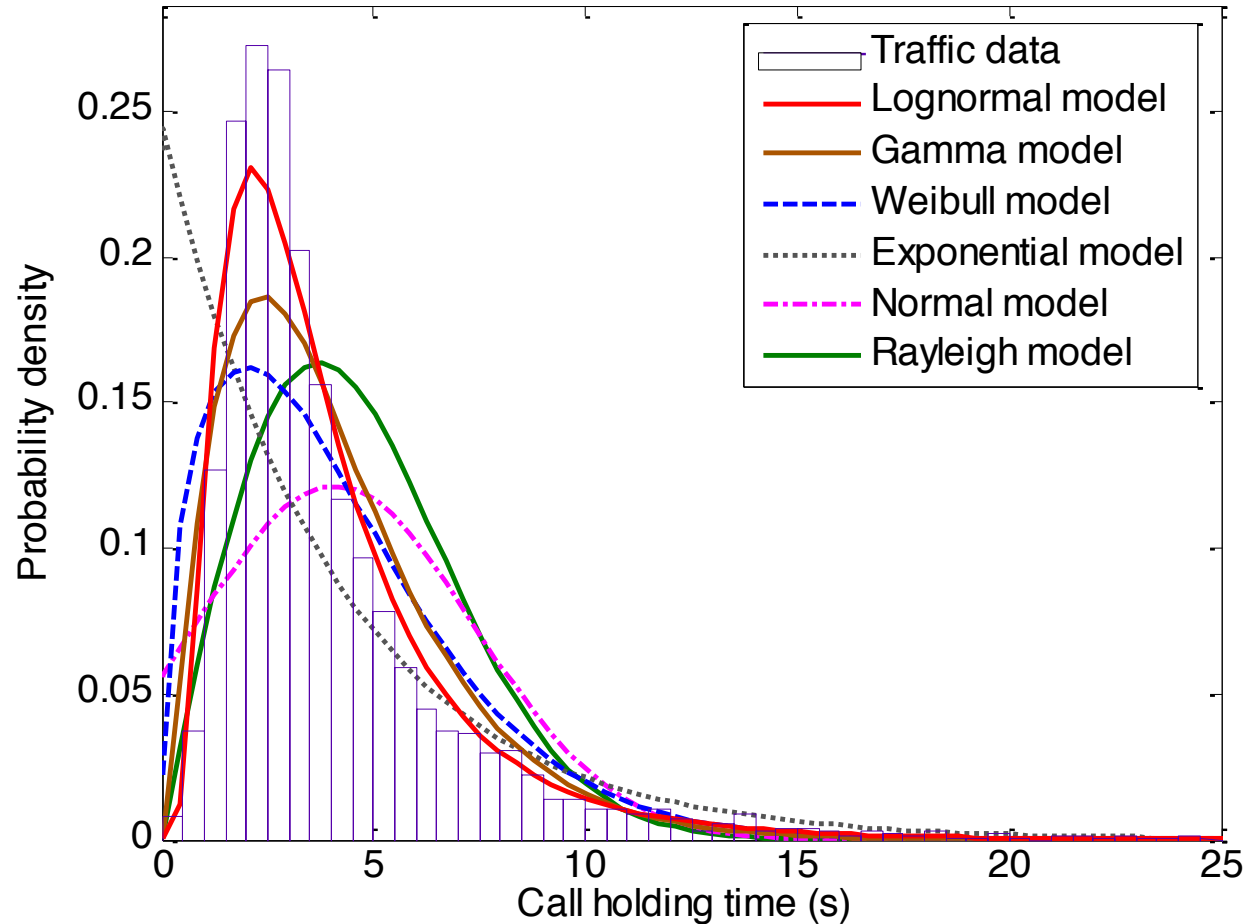


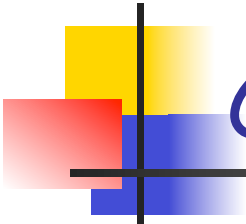
Call inter-arrival times: estimates of H

- Traces pass the test for time constancy of a :
estimates of H are reliable

2001		2002		2003	
Day/hour	H	Day/hour	H	Day/hour	H
02.11.2001 15:00–16:00	0.907	01.03.2002 04:00–05:00	0.679	26.03.2003 22:00–23:00	0.788
01.11.2001 00:00–01:00	0.802	01.03.2002 22:00–23:00	0.757	25.03.2003 23:00–24:00	0.832
02.11.2001 16:00–17:00	0.770	01.03.2002 23:00–24:00	0.780	26.03.2003 23:00–24:00	0.699
01.11.2001 19:00–20:00	0.774	01.03.2002 00:00–01:00	0.741	29.03.2003 02:00–03:00	0.696
02.11.2001 20:00–21:00	0.663	02.03.2002 00:00–01:00	0.747	29.03.2003 01:00–02:00	0.705

Call holding times: pdf candidates





Call holding times: estimates of H

- All (except one) traces pass the test for constancy of a
- only one unreliable estimate (*): consistent value

2001		2002		2003	
Day/hour	H	Day/hour	H	Day/hour	H
02.11.2001 15:00–16:00	0.493	01.03.2002 04:00–05:00	0.490	26.03.2003 22:00–23:00	0.483
01.11.2001 00:00–01:00	0.471	01.03.2002 22:00–23:00	0.460	25.03.2003 23:00–24:00	0.483
02.11.2001 16:00–17:00	0.462	01.03.2002 23:00–24:00	0.489	26.03.2003 23:00–24:00	0.463 *
01.11.2001 19:00–20:00	0.467	01.03.2002 00:00–01:00	0.508	29.03.2003 02:00–03:00	0.526
02.11.2001 20:00–21:00	0.479	02.03.2002 00:00–01:00	0.503	29.03.2003 01:00–02:00	0.466

Call inter-arrival and call holding times

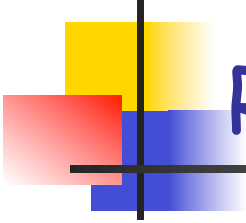
	2001		2002		2003	
	Day/hour	Avg. (s)	Day/hour	Avg. (s)	Day/hour	Avg. (s)
inter-arrival	02.11.2001 15:00–16:00	0.97	01.03.2002 04:00–05:00	0.81	26.03.2003 22:00–23:00	0.73
holding		3.78		4.07		4.08
inter-arrival	01.11.2001 00:00–01:00	0.97	01.03.2002 22:00–23:00	0.83	25.03.2003 23:00–24:00	0.85
holding		3.95		3.84		4.12
inter-arrival	02.11.2001 16:00–17:00	1.03	01.03.2002 23:00–24:00	0.86	26.03.2003 23:00–24:00	0.85
holding		3.99		3.88		4.04
inter-arrival	01.11.2001 19:00–20:00	1.09	01.03.2002 00:00–01:00	0.91	29.03.2003 02:00–03:00	0.87
holding		3.97		3.95		4.14
inter-arrival	02.11.2001 20:00–21:00	1.12	02.03.2002 00:00–01:00	0.91	29.03.2003 01:00–02:00	0.88
holding		3.84		4.06		4.25

Avg. call inter-arrival times: 1.08 s (2001), 0.86 s (2002), 0.84 s (2003)

Avg. call holding times: 3.91 s (2001), 3.96 s (2002), 4.13 s (2003)

Busy hour: best fitting distributions

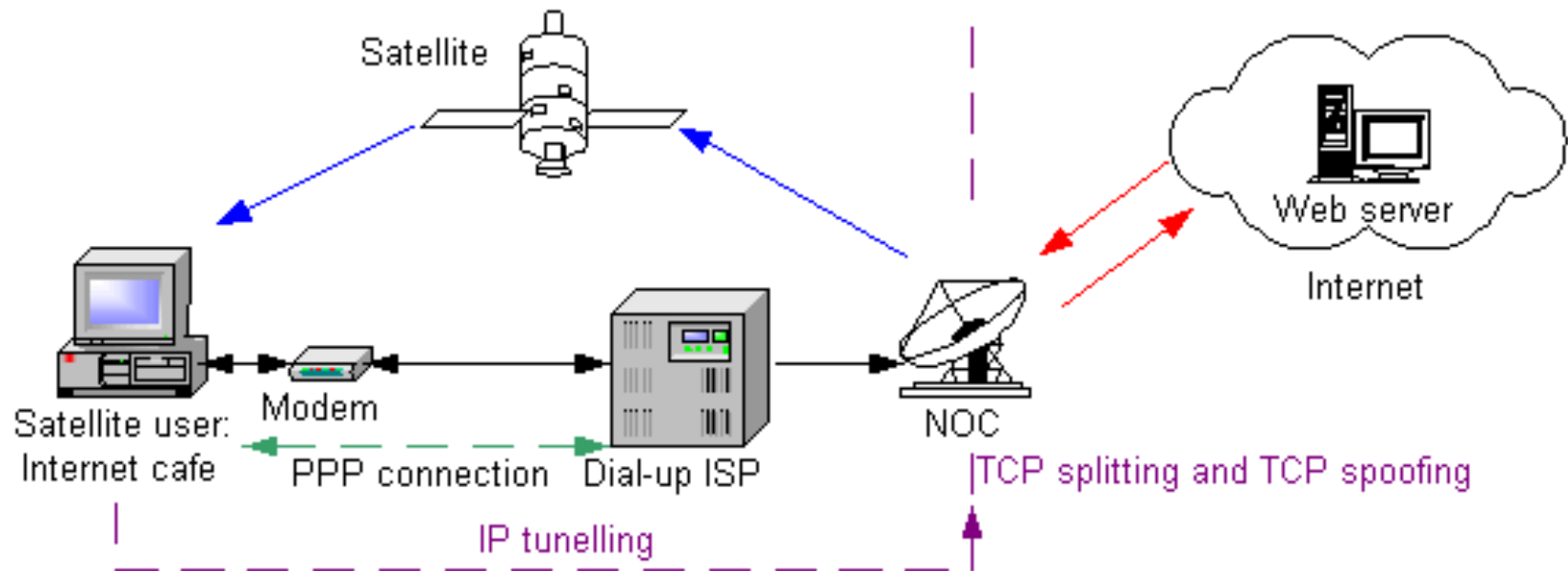
Busy hour	Distribution					
	Call inter-arrival times				Call holding times	
	Weibull		Gamma		Lognormal	
	a	b	a	b	μ	σ
02.11.2001 15:00–16:00	0.9785	1.1075	1.0326	0.9407	1.0913	0.6910
01.11.2001 00:00–01:00	0.9907	1.0517	1.0818	0.8977	1.0801	0.7535
02.11.2001 16:00–17:00	1.0651	1.0826	1.1189	0.9238	1.1432	0.6803
01.03.2002 04:00–05:00	0.8313	1.0603	1.1096	0.7319	1.1746	0.6671
01.03.2002 22:00–23:00	0.8532	1.0542	1.0931	0.7643	1.1157	0.6565
01.03.2002 23:00–24:00	0.8877	1.0790	1.1308	0.7623	1.1096	0.6803
26.03.2003 22:00–23:00	0.7475	1.0475	1.0910	0.6724	1.1838	0.6553
25.03.2003 23:00–24:00	0.8622	1.0376	1.0762	0.7891	1.1737	0.6715
26.03.2003 23:00–24:00	0.8579	1.0092	1.0299	0.8292	1.1704	0.6696

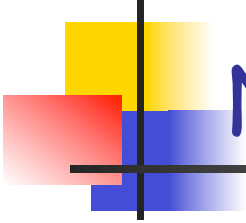


Roadmap

- 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

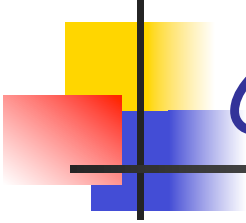
Case study: ChinaSat DirecPC system





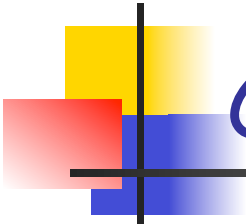
Network and traffic data

- **ChinaSat**: network architecture and TCP
- Analysis of **billing** records:
 - aggregated traffic
 - user behavior
- Analysis of **tcpdump** traces:
 - general characteristics
 - TCP options and operating system (OS) fingerprinting
 - network anomalies



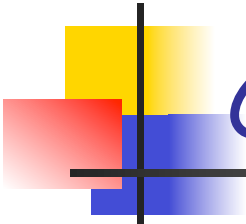
Characteristics of satellite links

- ChinaSat hybrid satellite network
 - Employs geosynchronous 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



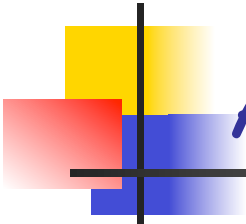
ChinaSat data: analysis

- ChinaSat traffic is self-similar and non-stationary
- **Hurst parameter** differs depending on traffic load
- Modeling of TCP connections:
 - inter-arrival time is best modeled by the **Weibull** distribution
 - number of downloaded bytes is best modeled by the **lognormal** distribution
- The distribution of visited websites is best modeled by the **discrete Gaussian exponential** (DGX) distribution



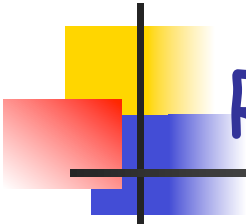
ChinaSat data: analysis

- Traffic prediction:
 - autoregressive integrative moving average (ARIMA) was successfully used to predict uploaded traffic (but not downloaded traffic)
 - wavelet + autoregressive model outperforms the ARIMA model
- Q. Shao and Lj. Trajkovic, "Measurement and analysis of traffic in a hybrid satellite-terrestrial network," *Proc. SPECTS 2004*, San Jose, CA, July 2004, pp. 329-336.



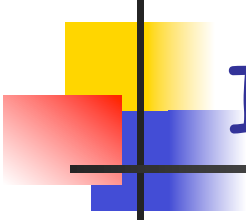
Analysis of collected data

- Analysis of patterns and statistical properties of two sets of data from the ChinaSat DirecPC network:
 - **billing** records
 - **tcpdump** traces
- **Billing** records:
 - daily and weekly traffic patterns
 - user classification:
 - single and multi-variable k-means clustering based on average traffic
 - hierarchical clustering based on user activity



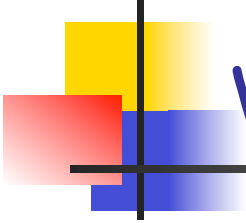
Roadmap

- 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



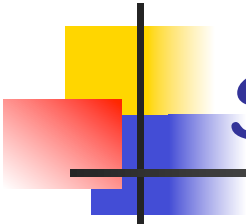
Internet topology

- 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)



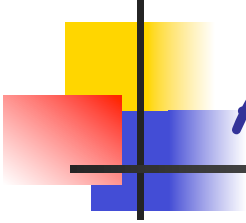
Variety of graphs

- **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



Analyzed datasets

- Sample datasets:

- Route Views:

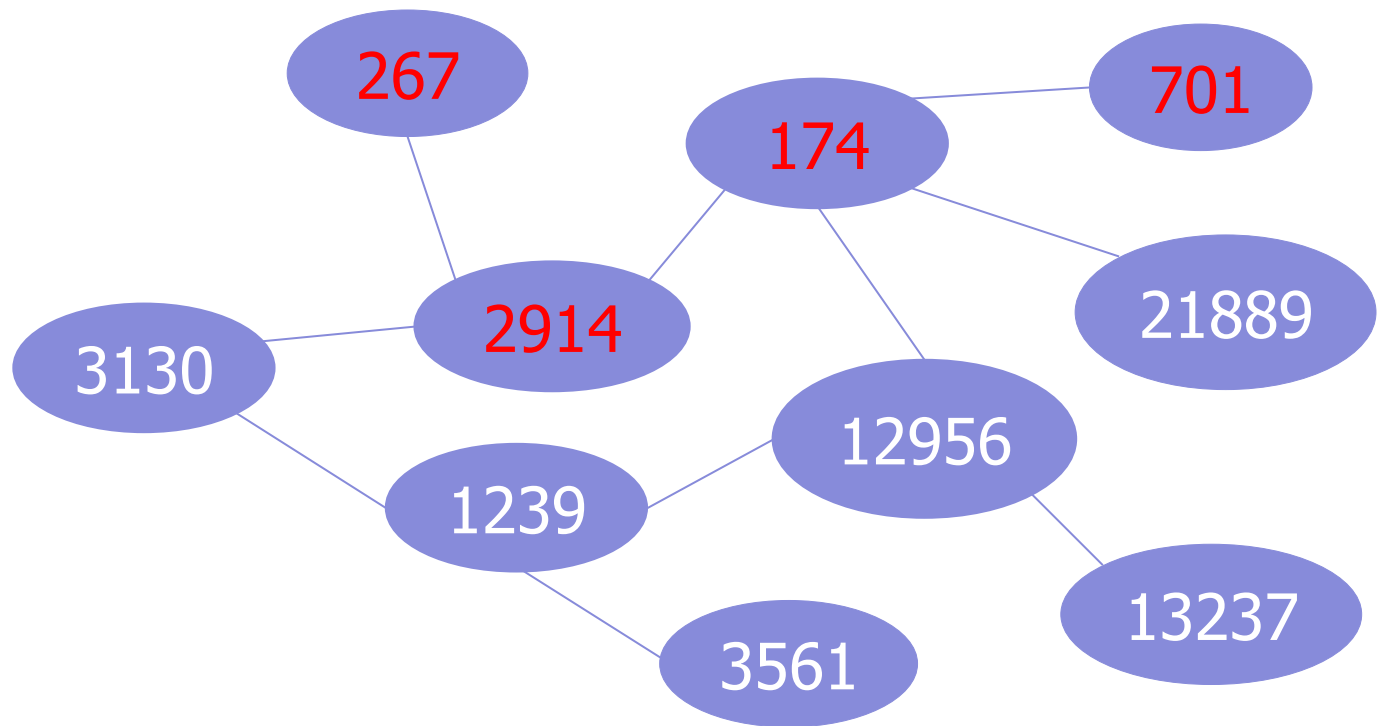
```
TABLE_DUMP| 1050122432| B| 204.42.253.253|  
267| 3.0.0.0/8| 267 2914 174 701| IGP|  
204.42.253.253| 0| 0| 267:2914 2914:420  
2914:2000 2914:3000| NAG| |
```

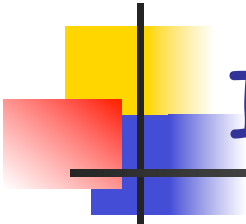
- RIPE:

```
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

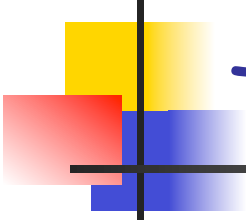
- Datasets collected from Border Gateway Protocols (BGP) routing tables are used to infer the Internet topology at AS-level





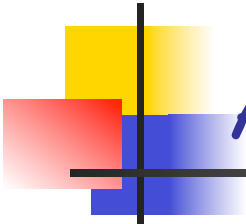
Internet topology

- 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



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

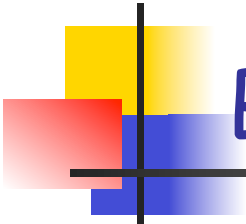


Anomaly detection techniques

- 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:

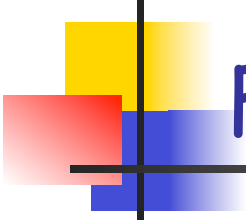
- Detect Internet routing anomalies using the Border Gateway Protocol (BGP) update messages



BGP features

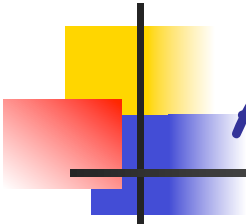
Approach:

- 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



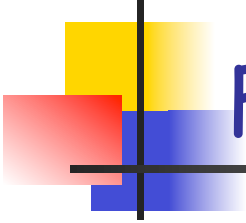
Feature selection

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



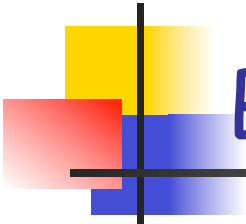
Anomaly classification

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



Feature extraction: BGP messages

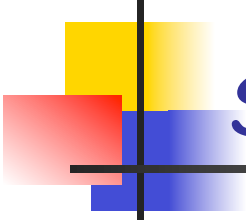
- Border Gateway Protocol (BGP) enables exchange of routing information between gateway routers using update messages
- BGP update message collections:
 - Réseaux IP Européens (RIPE) under the Routing Information Service (RIS) project
 - Route Views
 - Available in multi-threaded routing toolkit (MRT)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



BGP: known anomalies

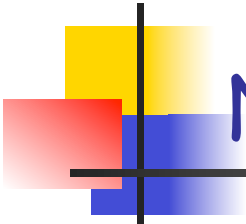
Anomaly	Date	Duration (h)
Slammer	January 25, 2003	16
Nimda	September 18, 2001	59
Code Red I	July 19, 2001	10

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



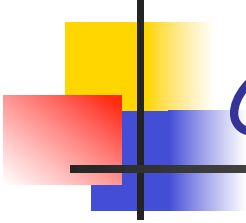
Slammer worm

- Sends its replica to randomly generated IP addresses
- Destination IP address gets infected if:
 - it is a Microsoft SQL serveror
 - a personal computer with the Microsoft SQL Server Data Engine (MSDE)



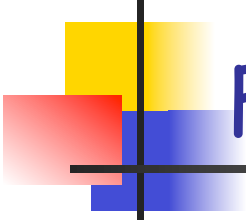
Nimda worm

- 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



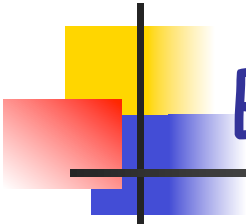
Code Red I worm

- 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



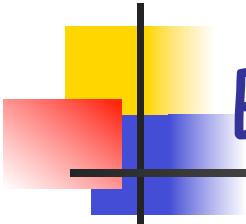
Feature extraction: BGP messages

- 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



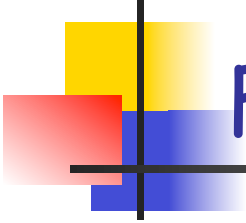
BGP features

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



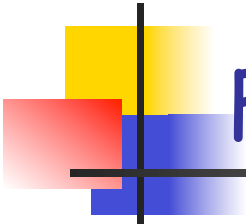
BGP features

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 $n = (7, \dots, 17)$	AS-path
25-33	Maximum AS-path length = n , where $n = (7, \dots, 15)$	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

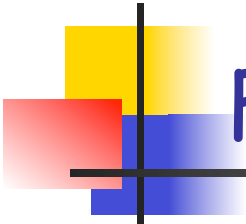
- Employed to select the most relevant features:
 - Fisher
 - Minimum Redundancy Maximum Relevance (mRMR)
 - Odds Ratio
 - Decision Tree
 - Fuzzy Rough Sets



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

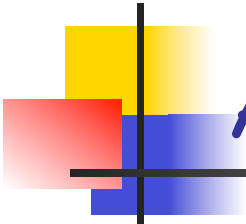
- Either four (30, 31, 32, 33) or five (22, 30, 31, 32, 33) features are removed in the constructed trees mainly because:
 - features are numerical and some are used repeatedly



Feature selection: fuzzy rough sets

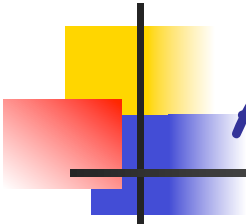
Dataset	Training data	Selected Features
Dataset 4	Slammer	1, 3-6, 9, 10, 13-32, 35
Dataset 5	Nimda	1, 3-4, 8-10, 12, 14-32, 35, 36
Dataset 6	Code Red I	3-4, 8-10, 12, 14-32, 35, 36

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



Anomaly classification

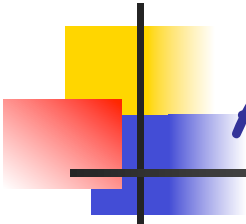
- Train classifiers for BGP anomaly detection using:
 - Support Vector Machines
 - Hidden Markov Models
 - Naive Bayes
 - Decision Tree
 - Extreme Learning Machine (ELM)



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

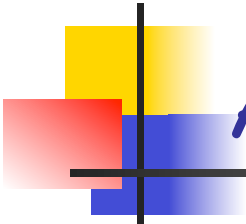
- Each path from the root node to a leaf node may be transformed into a decision rule
- A set of rules that are obtained from a trained decision tree may be used for classifying unseen samples



Anomaly classifiers: ELM

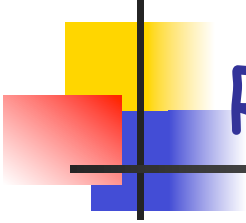
No. of features	Dataset	Acc_{train}	Acc_{test}	Training time (s)
37	Dataset 1	83.57 ± 0.11	80.01 ± 0.07	2.3043
	Dataset 2	83.53 ± 0.12	79.75 ± 0.08	2.2756
	Dataset 3	80.82 ± 0.09	21.65 ± 1.93	2.2747
17	Dataset 1	84.50 ± 0.07	79.91 ± 0.01	1.9268
	Dataset 2	84.43 ± 0.12	79.53 ± 0.10	1.5928
	Dataset 3	83.06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



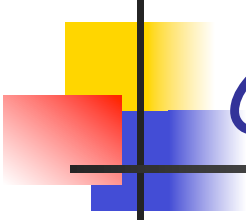
Anomaly classifiers: ELM

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



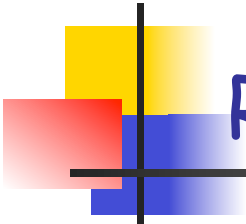
Roadmap

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



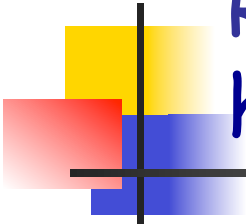
Conclusions

- 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/](http://www.routeviews.org/).
- CAIDA: Center for Applied Internet Data Analysis: [Online]. Available:
<http://www.caida.org/home/>.



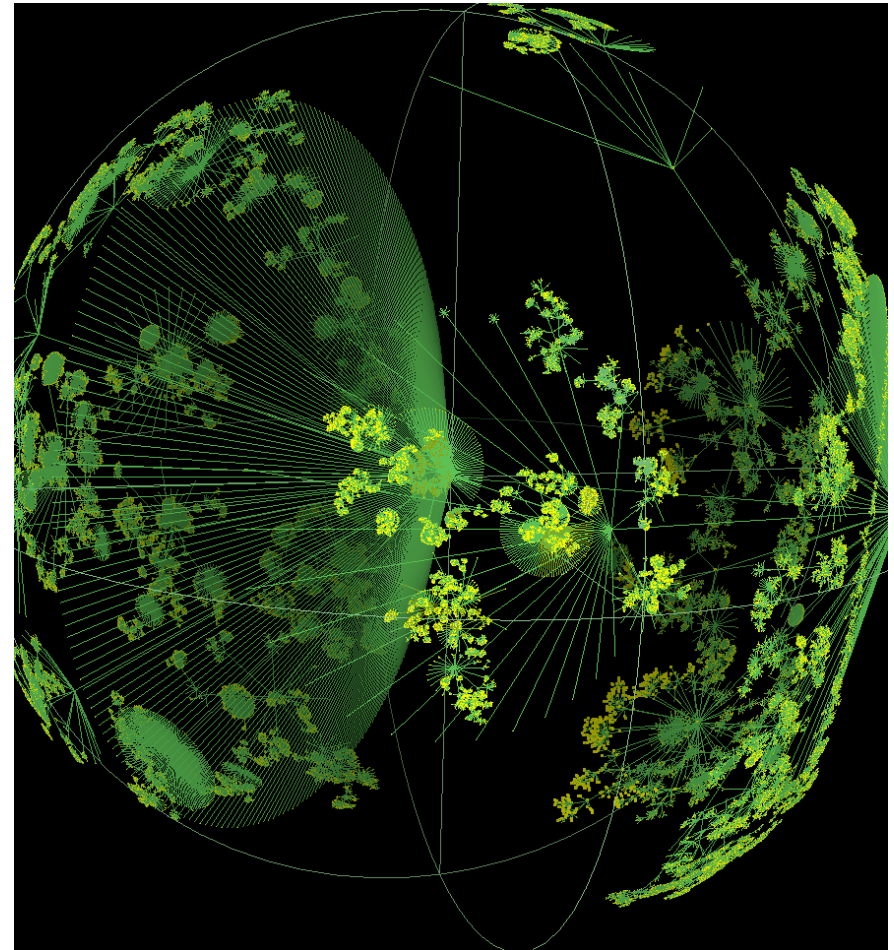
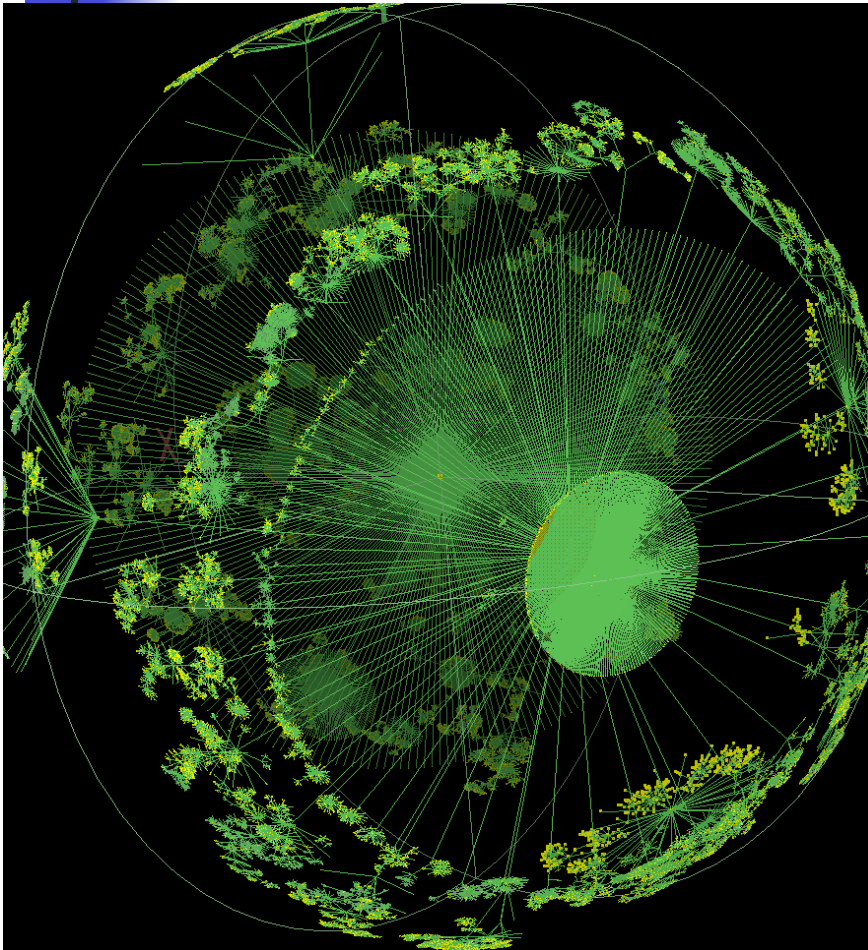
References:

<http://www.sfu.ca/~ljilja/cnl>

- 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.
- 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. 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.



Ihr: 535,102 nodes and 601,678 links



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