MODELING AND CHARACTERIZATION OF TRAFFIC IN DEPLOYED NETWORKS

Ljiljana Trajković
ljilja@cs.sfu.ca

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

Roadmap

- Introduction
- Traffic data and analysis tools:
  - data collection
  - statistical analysis
  - clustering tools
  - prediction analysis
- Two case studies:
  - satellite network: ChinaSat
  - public safety wireless network: E-Comm
- Conclusions and references

Collaborators

- M.A.Sc. and M.Eng. students:
  - Duncan Sharp
  - Nikola Cackov
  - Leo Chen
  - Bozidar Vujičić
  - Svetlana Vujičić
  - Nenad Lasković
  - Qing Shao

Relevant publications

- N. Cackov, J. Song, B. Vujičić, S. Vujičić, and Lj. Trajković, "Simulation of a public safety wireless networks: a case study," *Simulation*, SCS, to appear.
- 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.
- J. Song and Lj. Trajković, "Modeling and performance analysis of public safety wireless networks," in *Proc. IEEE IPCCC*, Phoenix, AZ, Apr. 2005, pp. 567–572.
- H. Chen and Lj. Trajković, "Trunked radio systems: traffic prediction based on user clusters," in *Proc. IEEE ISWCS 2004*, Mauritius, Sept. 2004, pp. 76–80.
- D. Sharp, N. Cackov, N. Lasković, Q. Shao, and Lj. Trajković, "Analysis of public safety traffic on trunked land mobile radio systems," *IEEE J. Select. Areas Commun.*, vol. 22, no. 7, pp. 1197–1205, Sept. 2004.
- Q. Shao and Lj. Trajković, "Measurement and analysis of traffic in a hybrid satellite-terrestrial network," in *Proc. SPECTS 2004*, San Jose, CA, July 2004, pp. 329–336.
- N. Cackov, B. Vujičić, S. Vujičić, and Lj. Trajković, "Using network activity data to model the utilization of a trunked radio system," in *Proc. SPECTS*, San Jose, CA, July 2004, pp. 517–524.

Network traffic measurements

- Focus of networking research during:
  - mid to late 1980's
  - early 1990's
- Motivation for traffic measurements:
  - understand traffic characteristics in deployed networks
  - develop traffic models
  - evaluate performance of protocols and applications
  - perform trace driven simulations

Case studies: deployed networks

- Analysis of traffic from operational networks:
  - provides useful information about the user behavior patterns
  - enables network operators to better understand the behavior of network users
  - helps provide better quality of service
- Traffic prediction: important to assess future network capacity requirements and to plan future network developments

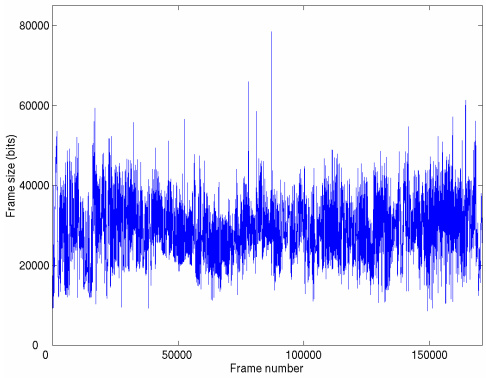
Roadmap

- Introduction
- Traffic data and analysis tools:
  - data collection
  - statistical analysis
  - clustering tools
  - prediction analysis
- Two case studies:
  - satellite network: ChinaSat
  - wireless network: E-Comm
- Conclusions and references

Traffic traces

- Most available traffic traces are from wired networks within research communities:
  - Bellcore, LBNL, Auckland University
- Few traces have been collected from wireless or satellite commercial networks
- Various factors affect Internet traffic patterns:
  - Web, Proxy, Napster, MP3, Web mail

Traffic patterns: "Star Wars" trace



- 170,000 frames (2 hours)
- 24 frames a second
- each source starts at a random point within the trace
- at the end, the trace wraps around to its beginning

Video traces

- MPEG-1 and MPEG-4 encoded video traces:
  - range: 22,498–89,998 frames per trace
  - coding: 25 frames/second
  - duration: 15–60 minutes of video
- Each sample represents the size of a frame in bits
- Traces:
  - University of Wuerzburg, index of MPEG traces: http://www-info3.informatik.uni-wuerzburg.de/MPEG/traces
  - University of Berlin, MPEG-4 video traces for network performance evaluation: http://www.tkn.ee.tu-berlin.de/research/trace/trace.html

Roadmap

- Introduction
- Traffic data and analysis tools:
  - data collection
  - statistical analysis
  - clustering tools
  - prediction analysis
- Two case studies:
  - satellite network: ChinaSat
  - wireless network: E-Comm
- Conclusions and references

Self-similarity

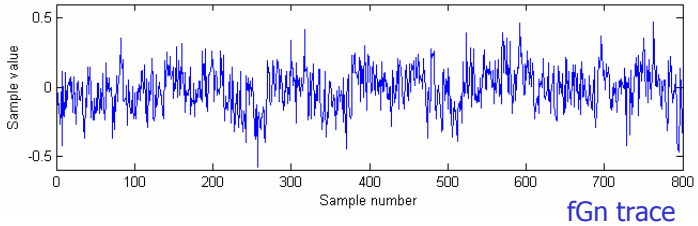
- Self-similarity implies a "fractal-like" behavior: data on various time scales have similar patterns
- A wide-sense stationary process X(n) is called (exactly second order) self-similar if its autocorrelation function satisfies:
  - $r^{(m)}(k) = r(k)$, $k \geqslant 0$, $m = 1, 2, \ldots, n$, where m is the level of aggregation
- Implications:
  - no natural length of bursts
  - bursts exist across many time scales
  - traffic does not become "smoother" when aggregated (unlike Poisson traffic)

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

Long-range dependence: properties

- High variability:
  - when the sample size increases, variance of the sample mean decays more slowly than expected
- Burstiness over a range of timescales:
  - long runs of large values followed by long runs of small values, repeated in aperiodic patterns



fGn trace

LRD: definition

- Slow decay of the autocorrelation function $r(k)$ of a (wide-sense) stationary process $X(n)$:

$$\sum_{k=-\infty}^{\infty} r(k) = \infty$$ definition

$$r(k) = c_r k^{-(2-2H)}, \ k \to \infty$$ model

$$f(\nu) = c_f \, |\nu|^{-\alpha}, \ \nu \to 0$$ corollary

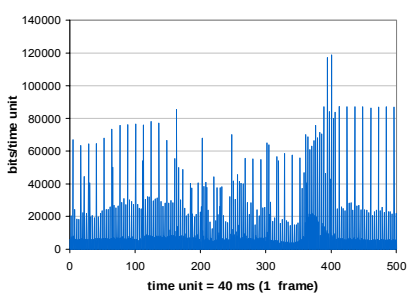
where $f(\nu)$ is the power spectral density of $X(n)$, c_r and c_f are non-zero constants, and $0<\alpha<1$

$0.5 < H < 1$ implies LRD

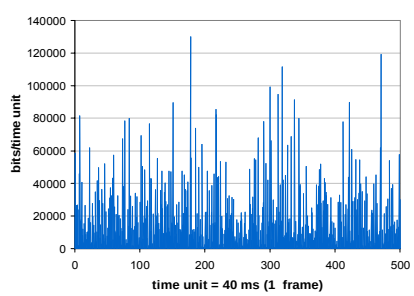
LRD: long-range dependence

Self-similar traffic patterns

Genuine MPEG traffic trace



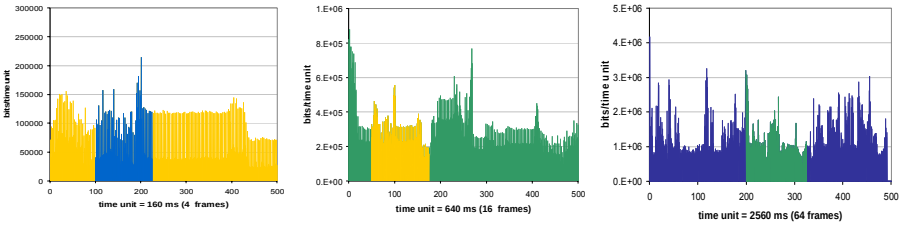
Poisson model



The two traces have identical mean.

Influence of time-scales

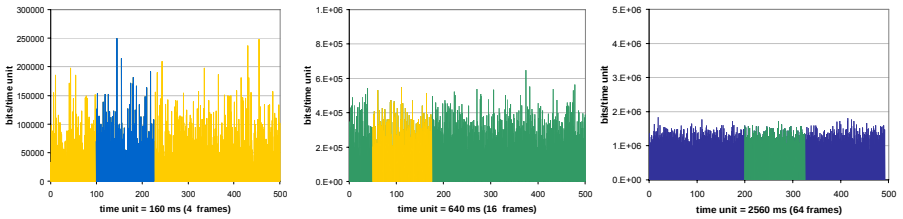
- Genuine MPEG traffic trace:



W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Trans. Networking,* vol. 2, pp. 1–15, 1994.

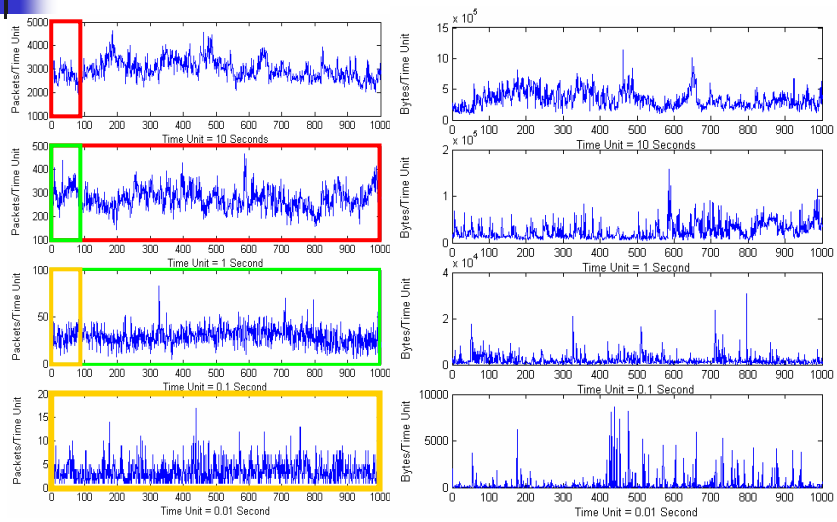
Influence of time-scales

- Synthetically generated Poisson model:

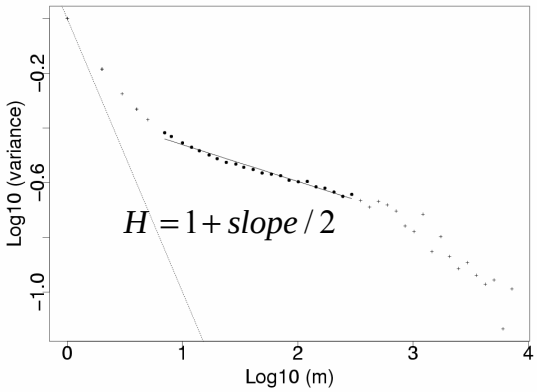


W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Trans. Networking,* vol. 2, pp. 1–15, 1994.

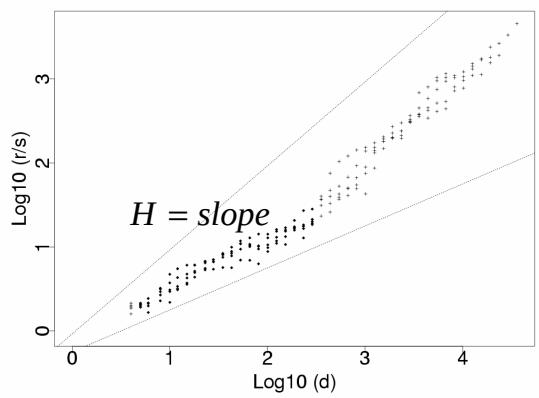
Self-similarity



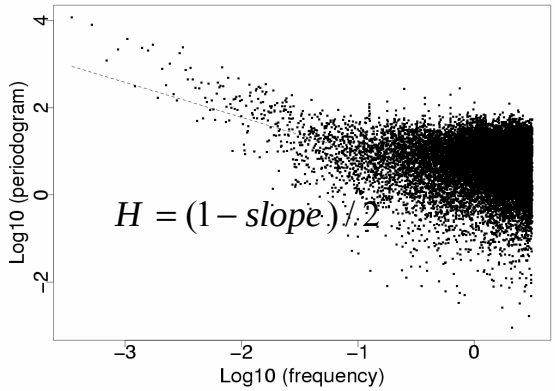
Estimation of H: variance-time plot



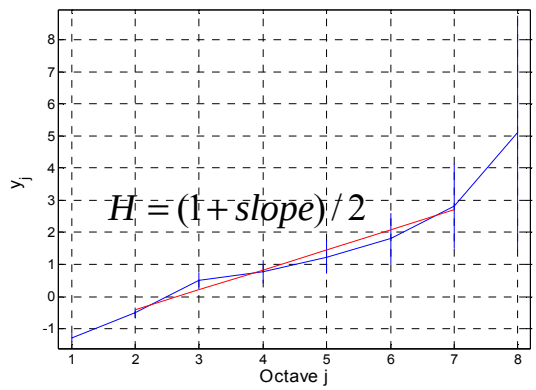
Estimation of H: R/S plot



Estimation of H: periodogram



Estimation of H: wavelets



Wavelet coefficients

- Discrete wavelet transform of a signal $X(t)$:

$$d(j,k) = \int_{-\infty}^{\infty} X(t)\psi_{j,k}(t)dt$$ wavelet coefficients

where

$$\psi_{j,k}(t) = 2^{-j/2}\psi\left(2^{-j}t - k\right)$$

- $\psi(t)$: mother wavelet
  - j : octave
  - k : translation
- Reconstruction formula:

$$X(t) = \sum_{j=0}^{\infty}\sum_{k} d(j,k)\psi_{j,k}(t)$$

LRD and wavelets

- Let $X(t)$ be LRD process (wide-sense stationary)
  - its power spectral density:

    $$f(\nu) \sim c_f \, |\nu|^{-\alpha}, \ \nu \to 0$$
- Mean square value of its wavelet coefficients on octave j satisfies:

  $$\mathrm{E}\{d(j,k)^2\} = 2^{j\alpha} c_f C(\alpha,\psi)$$

  where $C(\alpha,\psi) = \int |\nu|^{-\alpha} \left|\Psi(\nu)\right|^2 d\nu$ does not depend on j

D. Veitch and P. Abry, "A wavelet-based joint estimator of the parameters of long-range dependence," *IEEE Trans. on Information Theory*, vol. 45, no. 3, pp. 878–897, April 1999.

LRD and wavelets

- Logarithm:

  $$\log_2 \mathrm{E}\{d(j,k)^2\} = \alpha j + c$$
- Important property: for given j, $d(j,k)$ does not exhibit long-range dependence (with respect to k)
  - with appropriately chosen mother wavelet
- Hence:
  - simple estimator for $\mathrm{E}\{d(j,k)^2\}$ is a sample mean:

    $$\mathrm{E}\{d(j,k)^2\} = \frac{1}{n_j} \sum_{k=1}^{n_j} d(j,k)^2$$
  - n_j: number of wavelet coefficients at octave j

Estimation of α and H

- Logscale diagram: plot of $log_2E\{d(j,k)^2\}$ vs. j (octave)
- Linear relationship between $log_2E\{d(j,k)^2\}$ and j on the coarsest octaves indicates LRD
- Estimation of α:
  - linear regression of $log_2E\{d(j,k)^2\}$ on j in the linear region of the logscale diagram
- $H = 0.5\ (\alpha + 1)$

Roadmap

- Introduction
- Traffic data and analysis tools:
  - data collection
  - statistical analysis
  - clustering tools
  - prediction analysis
- Two case studies:
  - satellite network: ChinaSat
  - wireless network: E-Comm
- Conclusions and references

Clustering algorithms

- Clustering analysis groups or segments a collection of objects into subsets or clusters
- Objects within a cluster are more similar to each other than objects in distinct clusters
- An object can be described by a set of measurements or by its relations to other objects
- Network users are classified into clusters, according to the similarity of their behavior patterns
- We used AutoClass [Cheeseman and Stutz, 1996] and the K-means algorithm [Kaufman and Rousseeuw, 1990] to classify the calling patterns of talk groups

Clustering algorithms: AutoClass

- AutoClass is an unsupervised classification tool based on the Bayesian approach
- The goal of an unsupervised classification is to find the most probable set of class descriptions given the data and the prior expectations
- AutoClass begins by creating a random classification and then manipulates it into a high probability classification through local changes
- It repeats the process until it converges to a local maximum
- It starts over again and continues for a specified number of tries
- Each new try begins with a certain number of classes and may conclude with a smaller number of classes

Clustering algorithms: K-means

- The K-means algorithm is commonly used for data clustering
- Based on the input parameter k, it partitions a set of n objects into k clusters so that the resulting intra-cluster similarity is high and the inter-cluster similarity is low
- Similarity of clusters is measured with respect to the mean value of the objects in a cluster (viewed as the cluster's center of gravity)
- The algorithm is well-known for its simplicity and efficiency

Roadmap

- Introduction
- Traffic data and analysis tools:
  - data collection
  - statistical analysis
  - clustering tools
  - traffic prediction
- Two case studies:
  - satellite network: ChinaSat
  - wireless network: E-Comm
- Conclusions and references

Traffic prediction

- Traffic prediction: important to assess future network capacity requirements and to plan future network developments
- A network traffic trace consists of a series of observations in a dynamical system environment
- Traditional prediction: considers aggregate traffic and assumes a constant number of network users
- Approach that focuses on individual users has high computational cost for networks with thousands of users
- Employing clustering techniques for predicting aggregate network traffic bridges the gap between the two approaches

Traffic prediction: ARMA

- Predicting traffic using AutoRegressive Moving-Average (ARMA) model:

$$X(t) = \phi_1 X(t-1) + \cdots + \phi_p X(t-p) + e(t) + \theta_1 e(t-1) \cdots + \theta_q e(t-q)$$

$$\phi_P(B)\omega_t = \theta_q(B)\varepsilon_t$$

  where:
  - AR and MA parts: $\phi(B)$ and $\theta(B)$
  - B = the back-shift operator: $B^i X_t = X_{t-i}$
- Model:
  - past values: AutoRegressive (AR) structure
  - past random fluctuant effect: Moving Average (MA) process

Traffic prediction: ARIMA

- The general ARIMA model contains autoregressive (AR) and moving average (MA) parts and explicitly includes differencing in the formulation of the model
- Model parameters:
  - autoregressive parameter: p
  - number of differencing passes: d
  - moving average parameter: q
- ARIMA models are classified as ARIMA (p, d, q)

Traffic prediction: SARIMA

- Seasonal ARIMA is a variation of the ARIMA model
- It was applied to NSFNET traffic [Groschwitz and Polyzos, 1994] and sub-networks [Chen, 2002] for traffic model fitting and forecasting
- SARIMA additional model parameters:
  - seasonal period parameter: S
  - seasonal autoregressive parameter: P
  - number of seasonal differencing passes: D
  - seasonal moving average parameter: Q

Traffic prediction: SARIMA

- A SARIMA $(p,d,q)\times(P,D,Q)_S$ model can be represented as:

$$\phi(B^S)\phi(B)(1-B^S)^D(1-B)^d X_t = \theta(B^S)\theta(B)Z(t)$$

where:
  - AR and MA parts: $\phi(B)$ and $\theta(B)$
  - seasonal AR and seasonal MA parts: $\phi(B^s)$ and $\theta(B^s)$
  - B = the back-shift operator: $B^i X_t = X_{t-i}$

Roadmap

- Introduction
- Traffic data and analysis tools:
  - data collection
  - statistical analysis
  - clustering tools
  - prediction analysis
- Two case studies:
  - satellite network: ChinaSat
  - wireless network: E-Comm
- Conclusions and references

Roadmap

- Case study: ChinaSat
  - Introduction and motivation
  - Traffic:
    - collection
    - analysis
    - prediction

Case study: ChinaSat

- ChinaSat uses DirecPC (currently DirecWay) system to provide Internet access to over 200 Internet cafés across provinces:
  - satellite one-way broadcast system manufactured by Hughes Network Systems (deployed worldwide)
  - utilizes two special techniques to improve network performance:
    - IP spoofing
    - TCP splitting

Q. Shao and Lj. Trajković, "Measurement and analysis of traffic in a hybrid satellite-terrestrial network," in *Proc. SPECTS 2004*, San Jose, CA, July 2004, pp. 329–336.

DirecPC system

- IP spoofing:
  - customer's requests are not directly sent to the website
  - they are rerouted to the satellite network operation center (NOC)
  - NOC resends the requests to the website
  - website sends to the NOC data to be downloaded to the customer
- TCP splitting:
  - terrestrial links employ standard TCP
  - to improve throughput, space links with long delay use modified TCP versions with enlarged TCP window size

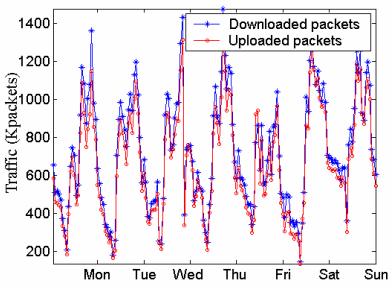
DirecPC system: traffic collection

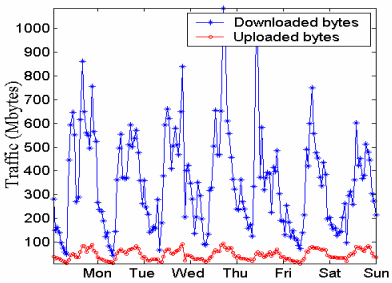
Red: uploaded traffic
Green: downloaded traffic

tcpdump trace format

- timestamp src > dst: flags data-seqno ack window urgent options
  - 19:12:45.660701 61.159.59.162.12800 > 192.168.1.169.62246: udp 52
  - 19:12:45.672959 192.168.1.242.40849 > 210.51.17.67.9065: P 6541284:6541321(37) ack 1479344110 win 8192 (DF)
  - 19:12:45.674709 192.168.2.30.39042 > 202.101.165.124.4220: . ack 807850998 win 8192
  - 19:12:45.676255 61.152.249.71.55901 > 192.168.1.242.40770: P 2627573783:2627573791(8) ack 5795719 win 63343 (DF)
  - 19:12:45.676256 61.152.249.71.55901 > 192.168.1.242.40846: P 2775973525:2775973533(8) ack 11622145 win 64102 (DF)
  - 19:12:45.688514 192.168.1.242.40770 > 61.152.249.71.55901: . ack 8 win 8192
  - 19:12:45.688843 192.168.1.242.40846 > 61.152.249.71.55901: . ack 8 win 8192
  - 19:12:45.689095 192.168.1.169.63644 > 202.103.69.103.3010: P 1969195:1969259(64) ack 2995916216 win 8192 (DF)
  - 19:12:45.692475 202.101.165.134.80 > 192.168.2.3.45585: . ack 3153903 win 6432
  - 19:12:45.699193 207.46.104.20.80 > 192.168.1.239.4912: R 2405276149:2405276149(0) win 0
- Red: uploaded traffic
- Green: downloaded traffic

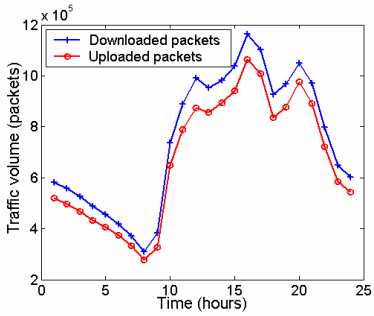
Analysis of weekly billing records

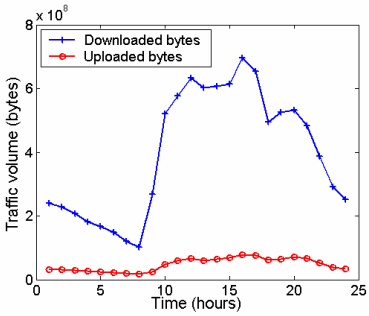




- Weekly traffic volume measured in packets (left) and bytes (right)
- Traffic data were collected from 09-12-2002 to 15-12-2002

Analysis of daily billing records





- Average traffic volume over a single day measured in packets (left) and bytes (right)
- Traffic data were collected from 9-12-2002 to 15-12-2002

Protocols and applications

Protocol	Packets	Packets (%)	Bytes	Bytes (%)
TCP	36,737,165	84.32	11,231,147,530	94.49
UDP	6,202,673	14.24	601,157,016	5.06
ICMP	630,528	1.45	53,128,377	0.45
Total	**43,570,366**	**~100**	**11,885,432,923**	**~100**

Applications	Connections	Connections (%)	Bytes	Bytes (%)
WWW	304,243	90.06	10,203,267,005	75.79
FTP-data	636	0.19	1,440,393,008	10.7
IRC	2,324	0.69	945,965	0.008
SMTP	562	0.17	2,326,373	0.01
POP-3	115	0.03	2,326,373	0.02
Telnet	70	0.02	280,286	0.002
Other	651	8.84	238,099,412	13.47
Total	**308,601**	**100**	**11,885,432,923**	**100**

- Traffic data were collected from 21-12-2002 22:08 to 23-12-2002 3:28

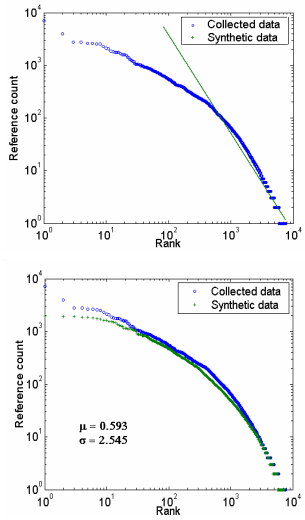
TCP connection level: Web traffic

- Zipf-like distribution:
  - the number of requests (frequency) is inversely proportional to its rank among the requests ($fr \sim 1/r^{\beta}$)
- DGX (discrete lognormal):

$$p(x=k)=\frac{A(\mu,\sigma)}{k}\exp[-\frac{(\ln k-\mu)^2}{2\sigma^2}]$$

$$A(\mu,\sigma)=\{\sum_{k=1}^{\infty}\frac{1}{k}[-\frac{(\ln k-\mu)^2}{2\sigma^2}]\}^{-1}$$

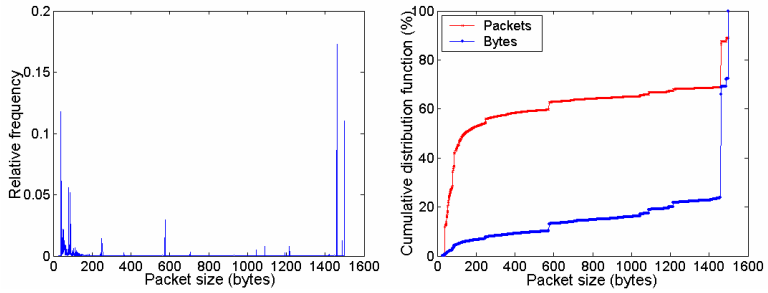
- DGX distribution fits better than the Zipf-like distribution



TCP connection level: Web traffic

- Traffic is non-uniformly distributed among the Internet hosts
- Ten busiest websites account for 60.23% of the entire traffic load:
  - all registered under the Asia Pacific Network Information Centre
  - the most popular site: a Chinese search engine website
- Language, geographical, and commercial factors (popular sites) greatly affect the traffic distribution
- Important for designing content delivery networks and caching proxies

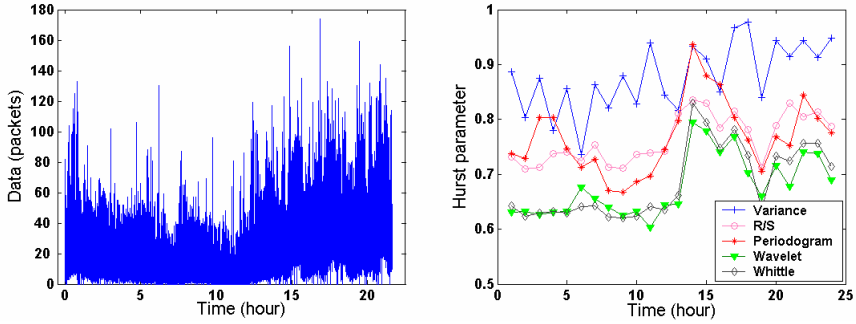
TCP packet size



Traffic data were collected from 21-12-2002 22:08 to 23-12-2002 3:28

- Packet size distribution is bimodal:
  - 50% of packets are smaller than 200 bytes
  - 30% of packets are greater than 1,400 bytes
- Most bytes are transferred in large packets

Estimation of self-similarity



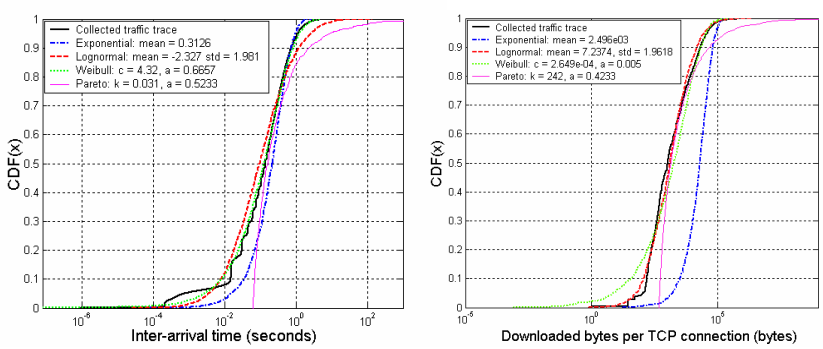
Traffic data were collected on 09-12-2002

TCP connection model

- We consider two parameters of a TCP connection:
  - connection inter-arrival times
  - number of downloaded bytes per connection
- Four probability distributions:

Distribution	Probability density	Cumulative probability
Exponential	$f(x)=\frac{1}{\rho}e^{-x/\rho}$	$F(x)=1-e^{-x/\rho}$
Weibull	$f(x)=\frac{1}{a}\left(-\frac{x}{a}\right)^{c-1}\mathrm{e}^{-(x/a)^c}$	$F(x)=1-e^{-(x/a)^c}$
Pareto $(k>0, a>0, x\geq k)$	$f(x)=\frac{ak^a}{(x)^{k+1}}$	$F(x)=1-\left(\frac{k}{x}\right)^a$
Lognormal	$f(x)=\frac{1}{x\sqrt{2\pi\sigma}}e^{-[\log(x)-\xi]^2/2\sigma^2}$	No closed form

TCP connection model



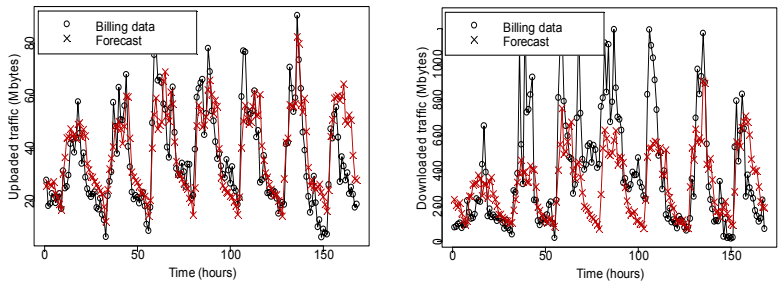
- Best fit:
  - Weibull: inter-arrival times of TCP connections
  - Lognormal: downloaded bytes per TCP connection

One-week ahead prediction

- Box-Jenkins method was applied to six weeks of billing records
- Derived parameters:
  - p=1, d = 0, q=0, P=0, D = 1, Q=1, S=168
  - collected traffic data fit the model $(1,0,0)\times(0,1,1)_{168}$
- Normalized mean squared error (nmse) is used to measure the performance of the predictor:

$$nmse = \frac{1}{\sigma^2 N}\sum_{k=1}^{N}(x(k)-\bar{x}(k))^2$$

Predictability evaluation



- Predicting downloaded traffic is more difficult than predicting uploaded traffic

Traffic type	Uploaded traffic	Downloaded traffic
nmse	0.3653	0.5988

ChinaSat case study: summary

- Analysis of collected traffic data:
  - Web applications and TCP protocol dominate the collected traffic
  - packet size distribution is bimodal and most bytes are transferred in large packets
  - few Web servers account for majority of data traffic
  - the frequency-rank relation of client connections matches the discrete lognormal distribution

ChinaSat case study: summary

- Self-similarity in collected traffic data:
  - various estimators of the Hurst parameter produced inconsistent results
  - more accurate estimation was achieved with the wavelet estimator
- TCP modeling:
  - Weibull: inter-arrival times of TCP connections
  - Lognormal: downloaded bytes per TCP connection
- Traffic prediction using the ARIMA model:
  - performs better for predicting the uploaded traffic
  - not suitable for predicting downloaded traffic

Roadmap

- Introduction
- Traffic data and analysis tools:
  - data collection
  - statistical analysis
  - clustering tools
  - prediction analysis
- Two case studies:
  - satellite network: ChinaSat
  - wireless network: E-Comm
- Conclusions and references

Roadmap

- Case study: E-Comm
  - Network utilization
  - Performance analysis
  - Modeling and characterization of traffic
  - Traffic prediction

Case study: wireless networks

- User behavior and mobility patterns in local-area and metropolitan-area wireless networks exhibit daily and weekly patterns [Tang and Baker, 1999]
- User behavior in Cellular Digital Packet Data (CDPD) mobile wireless networks has similar cyclic pattern [Andriantiatsaholiniaina and LT, 2002]
- Trunked radio network traffic [Sharp et al., 2004]:
  - the distribution of call holding times is approximately lognormal
  - call inter-arrival times are closely approximated by an exponential distribution

Case study: E-Comm network

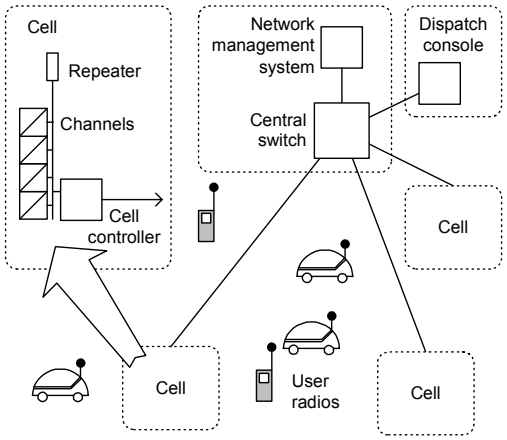
- E-Comm network: an operational trunked radio system serving as a regional emergency communication system
- The E-Comm network is capable of both voice and data transmissions
- Voice traffic accounts for over 99% of network traffic
- A group call is a standard call made in a trunked radio system
- More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

- Using network activity data to model the utilization of a trunked radio system
- Data and network models
- OPNET simulation results

N. Cackov, B. Vujičić, S. Vujičić, and Lj. Trajković, "Using network activity data to model the utilization of a trunked radio system," in *Proc. SPECTS*, San Jose, CA, July 2004, pp. 517–524.

Network architecture



Network characteristics

- Simulcast: all repeaters covering one cell use identical frequencies
- Trunking: all available frequencies in a cell are shared dynamically among all mobile users
- Cell capacity: number of available frequencies in a cell
  - one radio channel occupies one frequency
  - one call occupies one radio channel

cell	1	2	3	4	5	6	7	8	9	10	11
channels	12	7	4	5	3	7	6	4	6	6	3

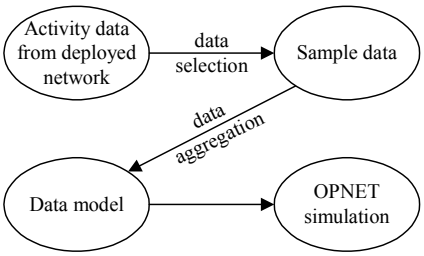
Call establishment

- Users are organized in talk groups:
  - one-to-many type of conversations
- Push-to-talk (PTT) mechanism for network access:
  - user presses the PTT button
  - system locates other members of the talk group
  - system checks for availability of channels:
    - channel available: call established
    - all channels busy: call queued/dropped
  - user releases PTT:
    - call terminates

Data processing

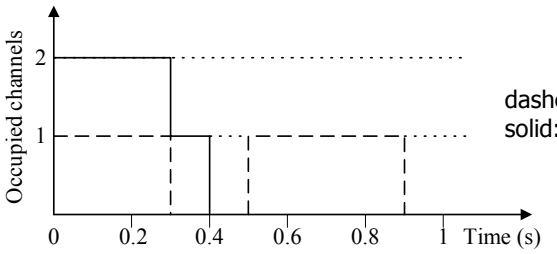
Timestamp	Duration (ms)	Caller	Callee	Cell
2003-03-20 0:00:10.639	4,870	A	B	4
2003-03-20 0:00:10.599	4,830	A	B	8
2003-03-20 0:00:10.529	4,860	A	B	9
2003-03-20 0:00:10.510	4,870	A	B	10

{10.510; 4,870; 4; 8; 9; 10}



Data discrepancies: 2002

- Coarse resolution of the timestamp
  - activity data: 10 ms
  - data model: 1 s
- Example:



dashed: deployed network
solid: model

Data discrepancies: 2003

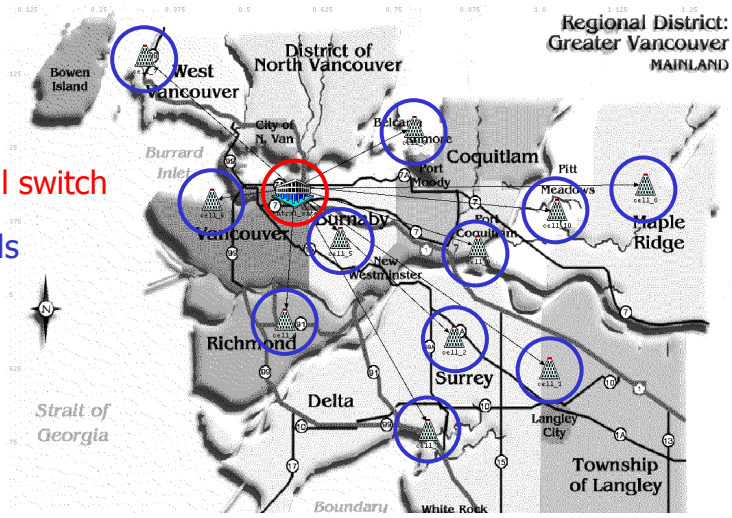
- Overlapping usage of channels

Timestamp	Duration (ms)	Cell	Channel
2003-03-20 0:00:33.370	9,420	10	4
…	…	…	…
2003-03-20 0:00:42.769	4,290	10	4

- 0:00:42.769 < 0:00:33.370 + 9.420
  - channel 4 in cell 10 is occupied by two calls at the same time!

Network model

- central switch
- 11 cells



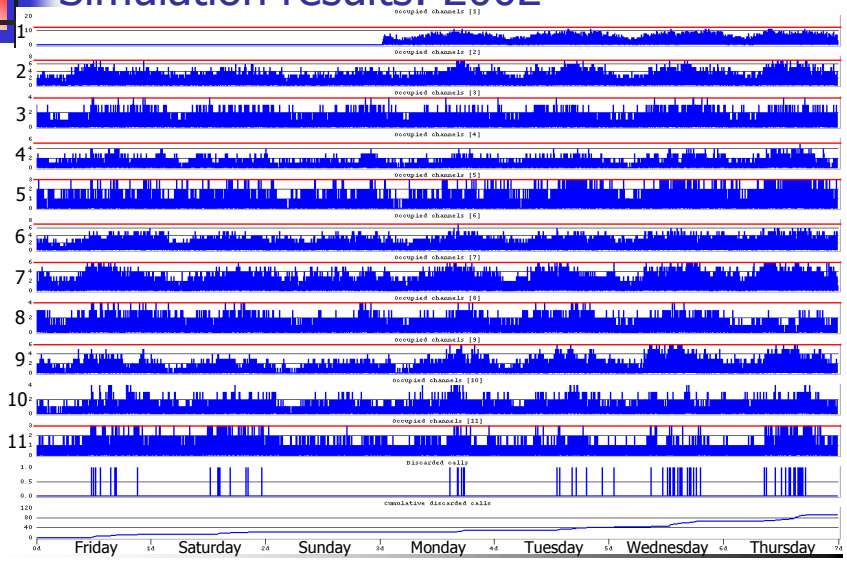
Network model: central switch

- Reads the trace file
- Generates packets according to the trace file
  - one call = one packet
  - packet_size (bits) = k × call_duration (s)
  - k: bit rate of channels (k=1,000 bps in simulations)
- Checks for availability of channels in the cells
- Collects statistics

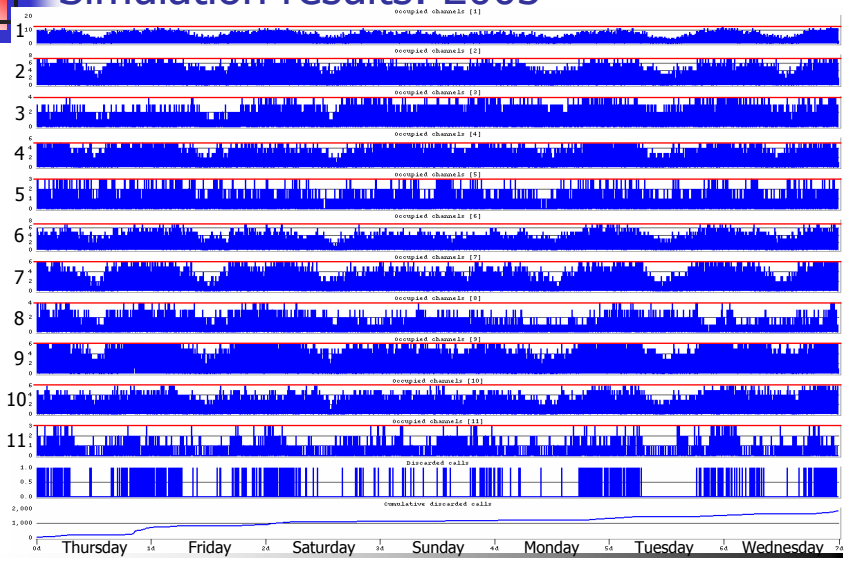
Data statistics

- 2002 sample data:
  - span: 8:00, February 1 – 8:00, February 8
  - number of calls: 403,590
  - discarded calls: 91
- 2003 sample data
  - span: 0:00, March 20 – 24:00, March 26
  - number of calls: 645,167
  - discarded calls: 1,812
- Discarded calls are due to discrepancies in the data
  - appear only in simulation results

Simulation results: 2002



Simulation results: 2003



Observations

- Presence of daily cycles:
  - minimum utilization: ~ 2 PM
  - maximum utilization: 9 PM - 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

Discarded calls

- appear only in the OPNET simulation results (do not exist in the deployed network)
- occur during busy hours
- can be used to identify possibly congested cells

Sample data	Cell no.	Capacity	No. of discarded calls
2002		original	91
2002	5	3 + 1	62
2003		original	1,812
2003	9	6 + 1	679
2003	4	5 + 1	521
	9	6 + 1	

original cap.	
cell	ch.
1	12
2	7
3	4
4	5
5	3
6	7
7	6
8	4
9	6
10	6
11	3

Maximum and average utilization

		2002		2003	
Cell	Capacity	Maximum	Average	Maximum	Average
1	12	11	2.5	11	2.6
2	7	7	0.8	7	1.6
3	4	4	0.3	4	0.5
4	5	5	0.3	5	1.1
5	3	3	0.2	3	0.3
6	7	7	0.7	7	1.2
7	6	6	0.7	6	1.1
8	4	4	0.3	4	0.4
9	6	6	0.4	6	1.6
10	6	4	0.2	6	1.0
11	3	3	0.2	3	0.2

Network utilization: summary

- We created a model in OPNET and simulated two weeks of network activity
- Network utilization exhibits daily cycles
- Between February 2002 and March 2003:
  - number of calls increased by ~ 60 %
  - average utilization increased non-uniformly across the network
- Several cells may become congested in future

Performance analysis

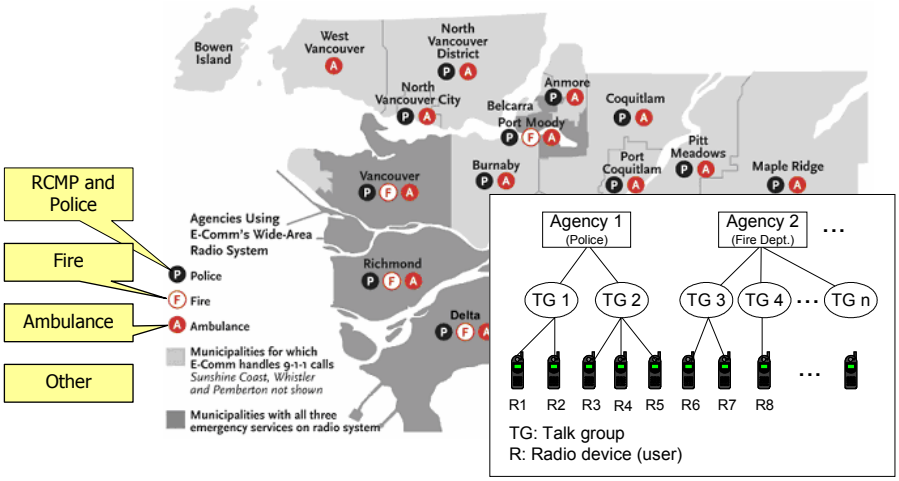
- Modeling and Performance Analysis of Public Safety Wireless Networks
- WarnSim: a simulator for public safety wireless networks (PSWN)
- Traffic data analysis
- Traffic modeling
- Simulation and prediction

J. Song and Lj. Trajković, "Modeling and performance analysis of public safety wireless networks," in *Proc. IEEE IPCCC*, Phoenix, AZ, Apr. 2005, pp. 567–572.

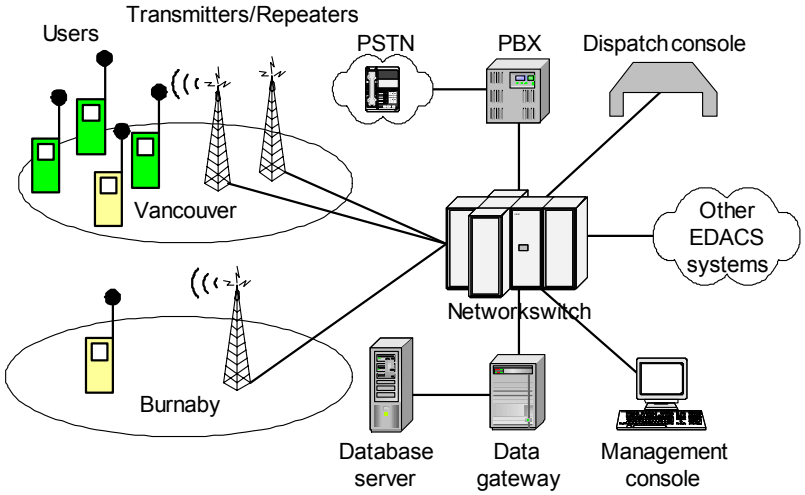
Motivations and objectives

- Motivations:
  - call traffic data from PSWN are rare
  - deployment costs of new channels are very high
- Objectives:
  - develop statistical models of call traffic
  - evaluate performance of the E-Comm network:
    - call blocking probability and channel utilization
  - Predict future network performance

E-Comm network: coverage and user agencies



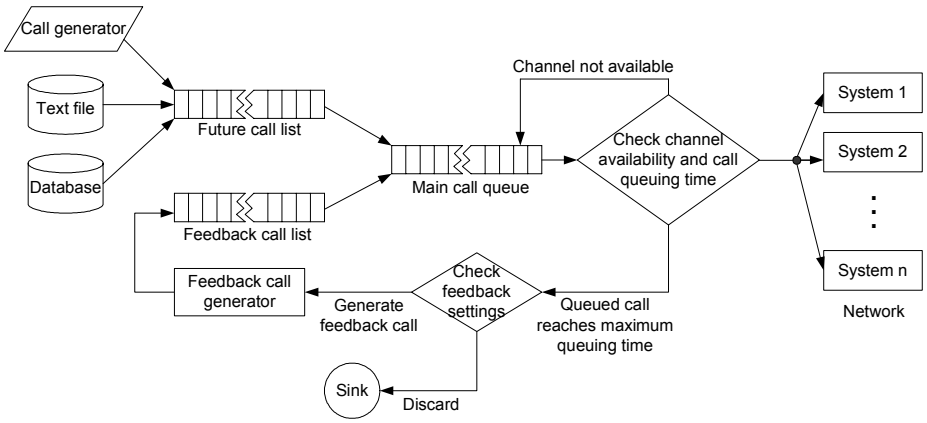
E-Comm network architecture



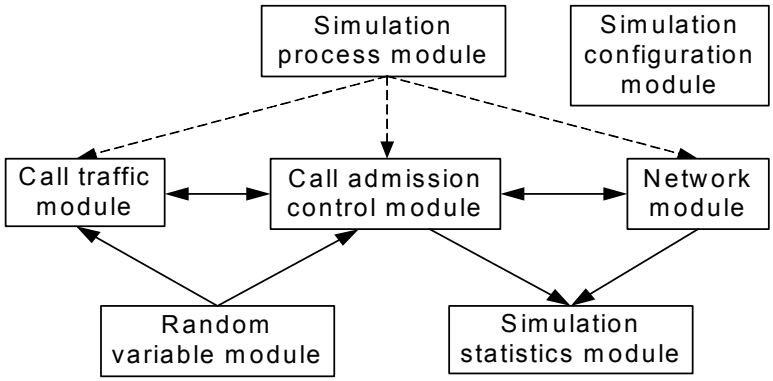
WarnSim overview

- Simulators such as OPNET, ns-2, and JSim are designed for packet-switched networks
- WarnSim is a simulator developed for circuit-switched networks, such as PSWN
- WarnSim:
  - publicly available simulator
    - http://www.vannet.ca/warnsim
  - effective, flexible, and easy to use
  - developed using Microsoft Visual C# .NET
  - operates on Windows platforms

WarnSim: high level diagram



WarnSim: module diagram



WarnSim modules (1)

- Network module:
  - models the PSWN cells (systems)
  - distributes calls to covered cells
  - tracks the number of free/occupied channels in each cell
  - periodically updates the channel status in each cell
- Call traffic module:
  - generates call traffic based on user-defined distributions
  - imports traffic trace from text files/databases for trace-driven simulations
  - combines traffic from multiple sources

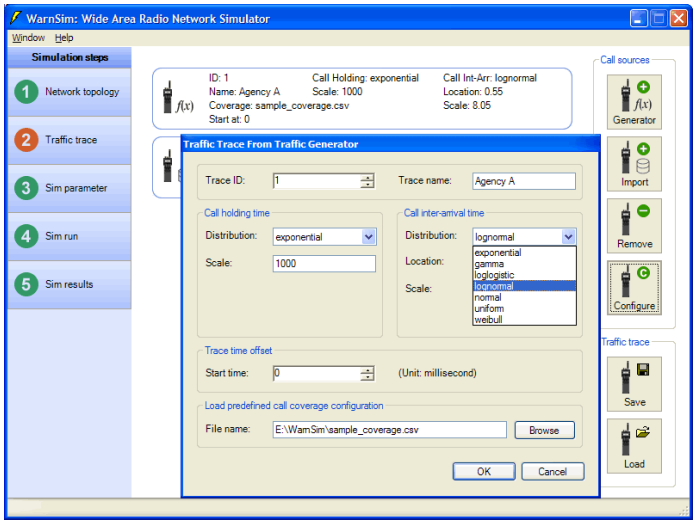
WarnSim modules (2)

- Call admission control module:
  - processes calls from the call traffic module
  - communicates with the network module to determine if there are available channels for a call to be established
  - manages the retrying mechanism of blocked calls
- Simulation configuration module:
  - keeps track of parameters such as call queuing mechanism and parameters such as maximum call queuing time, simulation duration, and simulation granularity

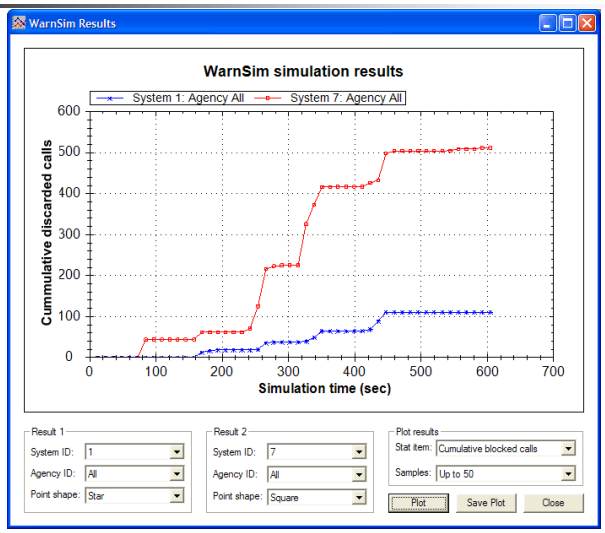
WarnSim modules (3)

- Simulation process module:
  - uses a timer to control and synchronize the operation of the WarnSim modules
- Simulation statistics module:
  - collects real-time and summary statistics of cells: number of calls, blocked calls, call blocking probability, and channel utilization
  - displays and visualizes simulation results
- Random variable module:
  - generates random variables: uniform, exponential, gamma, normal, lognormal, loglogistic, and Weibull

Traffic trace generator



WarnSim results



Traffic data

- 2 days of traffic data from 2001
- 28 days of traffic data from 2002
- 92 days of traffic data from 2003
- contain more than 10 million calls
- a sample record of the call traffic data:

Call arrival time	Duration (ms)	Caller agency
2003-05-01 00:00:09.620	1990	5

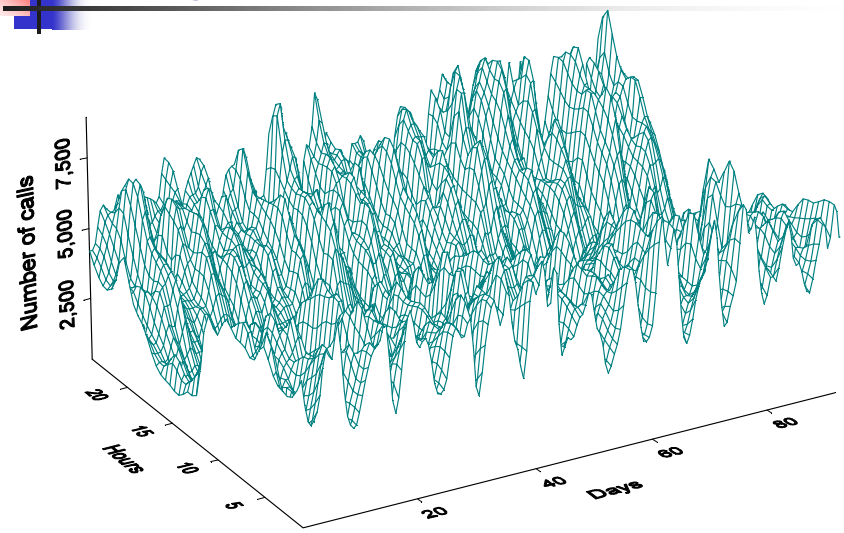
Caller	Callee	System ID	Channel no.
9999	1111	1, 7	3, 4

Call coverage

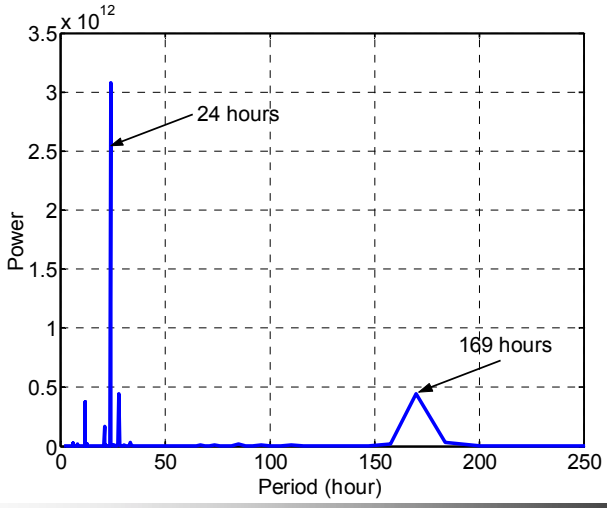
- Calls in the E-Comm network are group calls and most of them cover more than one system
- Call coverage pattern (systems a call covers) is determined by the organization of talk groups and deployment of radio devices
- Assumption: call coverage pattern remains constant

Call coverage	Percentage
System ID: 1, 2, 3	1.35%
System ID: 1, 3, 8, 10, 11	2.07%
System ID: 3, 5, 7, 9, 10	7.08%
…	…

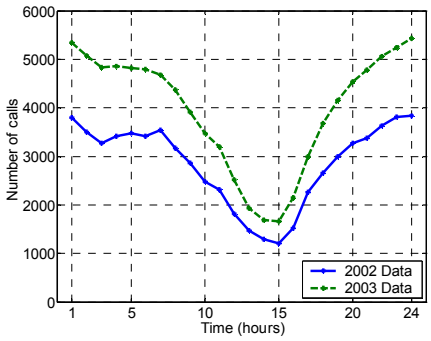
Hourly call arrival rate in 2003

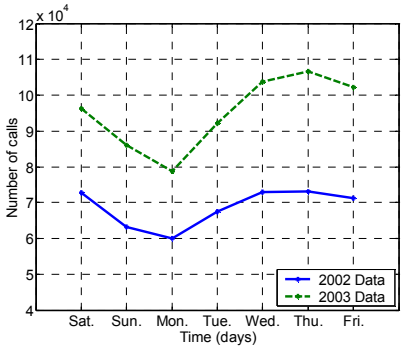


Hourly call arrival rate in 2003: power spectrum



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

Mobility of radio devices

- Mobility of radio devices and call handover are major concerns for micro-cell cellular networks
- They are of little importance in the E-Comm network:
  - E-Comm network is a wide area radio network with each system covering a citywide area
  - average call duration is 3.8 seconds
  - negligible probability for a radio device to move between two systems during such a short time

Traffic modeling procedure

- Modeling of traffic on user agency level
- Extract 500 data samples (call holding time/call inter-arrival time) from traffic data table
- Select a candidate distribution (exponential, lognormal, or gamma) and use Maximum Likelihood Estimation to estimate its parameters
- Use Kolmogorov-Smirnov goodness-of-fit (GoF) test to evaluate the candidate distribution

Candidate probability distributions

Exponential: $f(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}}$ mean

Gamma: $f(x) = \beta^{-k} \frac{x^{k-1}}{(k-1)!} e^{-\frac{x}{\beta}}$ shape scale

Lognormal: $f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{\frac{(\ln(x)-\mu)^2}{-2\sigma^2}}$ scale shape

Modeling call holding time: Agency 2

- Estimate distribution parameters and Kolmogorov-Smirnov GoF test results (p-value) for Agency 2 during the three busiest hours in 2002 dataset
- Lognormal distribution is likely to be accepted

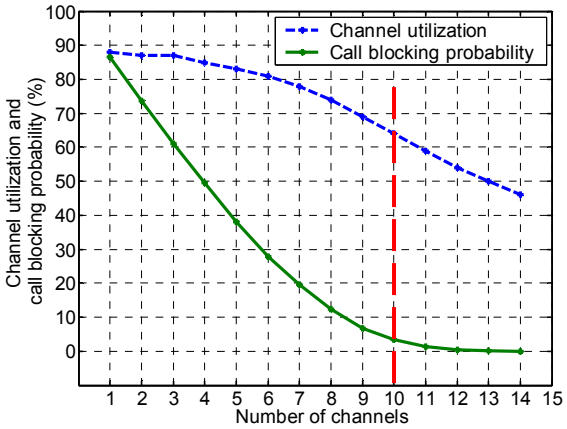
Distributions	Distribution parameters and K-S test results		
	Busy hour 1	Busy hour 2	Busy hour 3
Exponential	β = 3683.6 p-value = 0	β = 3664.06 p-value = 0	β = 3658.56 p-value = 0
Lognormal	σ = 8.0538 μ = 0.5505 p-value = 0.0832	σ = 8.0378 μ = 0.5769 p-value = 0.4952	σ = 7.9872 μ = 0.6334 p-value = 0.2634
Gamma	β = 1108.0657 k = 3.3246 p-value = 0.0002	β = 1173.6163 k = 3.1222 p-value = 0.0137	β = 1493.315 k = 2.4501 p-value = 0.0027

Proposed call traffic model

- Use lognormal distribution to model call holding time at user agency level
- Use exponential distribution to model call inter-arrival time at user agency level
- Assume call coverage pattern remains constant

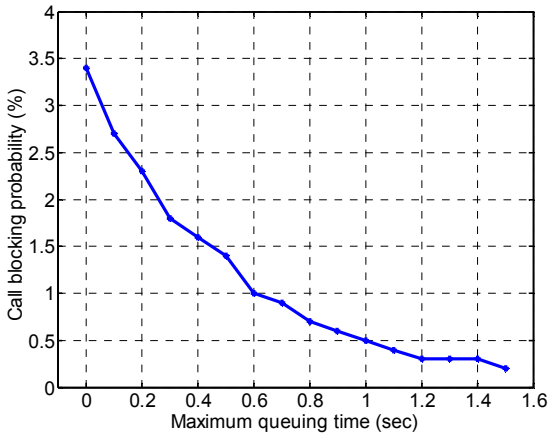
	Agency 2	Agency 5	Others
Call holding time	lognormal σ = 8.05 μ = 0.55	lognormal σ = 8.09 μ = 0.73	lognormal σ = 7.88 μ = 0.82
Call inter-arrival time	exponential β_1	exponential β_2	exponential β_3

Channels and Grade of Service: busy hour 2003-05-15 2:00–3:00 am



System ID	1	2	3	4	5	6	7	8	9	10	11
Channels	variable	7	4	5	3	7	8	4	7	6	3

Queuing and Grade of Service: busy hour 2003-05-15 2:00–3:00 am



System ID	1	2	3	4	5	6	7	8	9	10	11
Channels	10	7	4	5	3	7	8	4	7	6	3

Performance prediction

- Increased the number of calls made by Agency 5 by 100% from the busiest hour in 2003:
  - maximum queuing time is set to zero
  - parameters for WarnSim call traffic generator

	Agency 2	Agency 5	Others
Call holding time	lognormal $\sigma = 8.05$ $\mu = 0.55$	lognormal $\sigma = 8.09$ $\mu = 0.73$	lognormal $\sigma = 7.88$ $\mu = 0.82$
Call inter-arrival time	exponential $\beta = 1354$	exponential $\beta = 381$	exponential $\beta = 3480$

Performance prediction

System ID	Number of channels	Original blocking probability (%)	Original channel utilization (%)	Predicted blocking probability (%)	Predicted channel utilization (%)
1	10	1.9 – 3.5	57 – 65	12.1 – 12.6	71 – 72
2	7	0.0 – 0.6	29 – 48	5.1 – 7.0	54 – 55
3	4	0.0	11 – 14	0.1 – 0.4	16 – 17
4	5	0.0 – 0.4	21 – 23	1.5 – 3.7	35 – 39
5	3	0.0	4 – 17	1.0 – 1.3	16 – 18
6	7	0.0 – 0.3	19 – 42	2.1 – 2.7	44 – 45
7	8	0.0 – 0.4	25 – 34	0.6 – 0.8	38 – 40
8	4	0.0	8 – 11	0.0 – 0.3	16 – 18
9	7	0.3 – 0.5	37 – 43	9.0 – 10.1	60 – 62
10	6	0.0	16 – 26	1.3 – 1.5	35 – 38
11	3	0.0	6 – 10	0.3 – 0.9	11 – 13

Performance prediction

Additional channels:

+ 4 →
+ 2 →
+ 1 →
+ 3 →

for call blocking probability lower than 3%

System ID	Original blocking probability (%)	Original channel utilization (%)	Predicted blocking probability (%)	Predicted channel utilization (%)
1	1.9 – 3.5	57 – 65	12.1 – 12.6	71 – 72
2	0.0 – 0.6	29 – 48	5.1 – 7.0	54 – 55
3	0.0	11 – 14	0.1 – 0.4	16 – 17
4	0.0 – 0.4	21 – 23	1.5 – 3.7	35 – 39
5	0.0	4 – 17	1.0 – 1.3	16 – 18
6	0.0 – 0.3	19 – 42	2.1 – 2.7	44 – 45
7	0.0 – 0.4	25 – 34	0.6 – 0.8	38 – 40
8	0.0	8 – 11	0.0 – 0.3	16 – 18
9	0.3 – 0.5	37 – 43	9.0 – 10.1	60 – 62
10	0.0	16 – 26	1.3 – 1.5	35 – 38
11	0.0	6 – 10	0.3 – 0.9	11 – 13

Performance prediction

System ID	Number of channels	Original blocking probability (%)	Original channel utilization (%)	Predicted blocking probability (%)
1	10 + 4	1.9 – 3.5	57 – 65	< 3
2	7 + 2	0.0 – 0.6	29 – 48	< 3
3	4	0.0	11 – 14	0.1 – 0.4
4	5 + 1	0.0 – 0.4	21 – 23	< 3
5	3	0.0	4 – 17	1.0 – 1.3
6	7	0.0 – 0.3	19 – 42	2.1 – 2.7
7	8	0.0 – 0.4	25 – 34	0.6 – 0.8
8	4	0.0	8 – 11	0.0 – 0.3
9	7 + 3	0.3 – 0.5	37 – 43	< 3
10	6	0.0	16 – 26	1.3 – 1.5
11	3	0.0	6 – 10	0.3 – 0.9

Performance analysis: summary

- Developed a new tool (WarnSim) for simulating Public Safety Wireless Networks
- Analyzed call traffic data and showed the cyclic patterns of the call arrival rate
- Proposed statistical models for call traffic on user agency level
  - call holding time: lognormal distribution
  - call inter-arrival time: exponential distribution
- Using WarnSim:
  - evaluated the E-Comm network performance
  - predicted the future performance of the E-Comm network

Modeling and characterization of traffic

- Statistical concepts and analysis tools
- Analysis of traffic data:
  - call inter-arrival times
  - call holding times
- Traffic modeling and characterization

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.

Network characteristics

- EDACS: Enhanced Digital Access Communications Systems
- Simulcast: repeaters covering one cell use identical frequencies
- Trunking: available frequencies in a cell are shared dynamically among mobile users
- Cell capacity (number of available frequencies in a cell):
  - one radio channel occupies one frequency
  - one call occupies one radio channel

Call establishment

- Users are organized in talk groups:
  - one-to-many type of conversations
- Push-to-talk (PTT) mechanism for network access:
  - user presses the PTT button
  - system locates other members of the talk group
  - system checks for availability of channels:
    - channel available: call established
    - all channels busy: call queued/dropped
  - user releases PTT:
    - call terminates

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

Erlang models

- Erlang B model assumes:
  - call holding time follows exponential distribution
  - blocked call will be rejected immediately
- Erlang C model assumes:
  - call holding time follows exponential distribution
  - blocked call will be put into a FIFO queue with infinite size

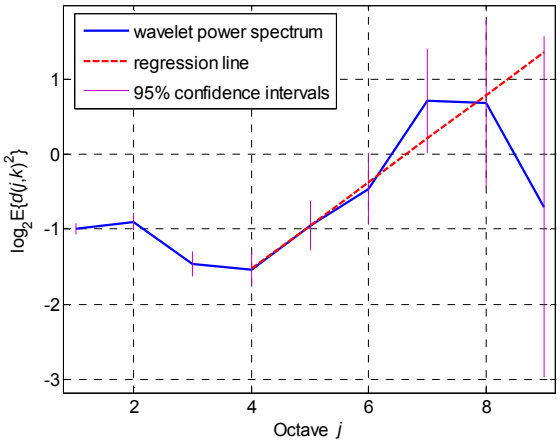
Statistical concepts

- Probability distribution:
  - probability that outcomes of a process are within a given range of values
  - expressed through probability density (pdf) and cumulative distribution (cdf) functions
- Autocorrelation:
  - measures the dependence between two outcomes of a process
  - wide-sense stationary processes: autocorrelation depends only on the difference (lag) between the time instances of the outcomes

Estimation of α and H

- Logscale diagram: plot of $\log_2 E\{d(j,k)^2\}$ vs. j (octave)
- Linear relationship between $\log_2 E\{d(j,k)^2\}$ and j on the coarsest octaves indicates LRD
- Estimation of α:
  - linear regression of $\log_2 E\{d(j,k)^2\}$ on j in the linear region of the logscale diagram
- $H = 0.5\ (\alpha + 1)$

Logscale diagram: example

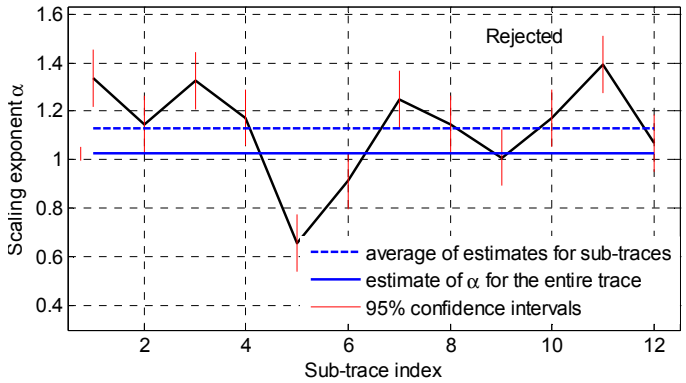


- call inter-arrival times: 22:00–23:00, 26.03.2003
- α=0.576, H=0.788 (octaves 4–9)

Test for time constancy of α

- $X(n)$: wide-sense stationary process
  - α does not depend on n
- Is α constant throughout the time series $X(n)$?
- Approach:
  - divide $X(n)$ into m blocks of equal lengths
  - estimate α for each block
  - compare the estimates
- If α varies significantly, estimating α for the entire time series is not meaningful
- In our analysis, m$\in${3, 4, 5, 6, 7, 8, 10}

Test for constancy: example



- Trace is divided into 12 sub-traces of equal lengths
- Variation of the scaling exponent indicates that α is not constant

Star Wars IV (MPEG-4)

Kolmogorov-Smirnov test

- Goodness-of-fit test: quantitative decision whether the empirical cumulative distribution function (ECDF) of a set of observations is consistent with a random sample from an assumed theoretical distribution
- ECDF is a step function (step size 1/N) of N ordered data points $Y_1, Y_2, \ldots, Y_N$:

$$E_N = \frac{n(i)}{N}$$

$n(i)$: the number of data samples with values smaller than Y_i

Parameters

- Hypothesis:
  - null: the candidate distribution fits the empirical data
  - alternative: the candidate distribution does not fit the empirical data
- Input parameters: significance level σ and tail
- Output parameters:
  - p-value
  - k: test statistic
  - cv: critical (cut-off) value

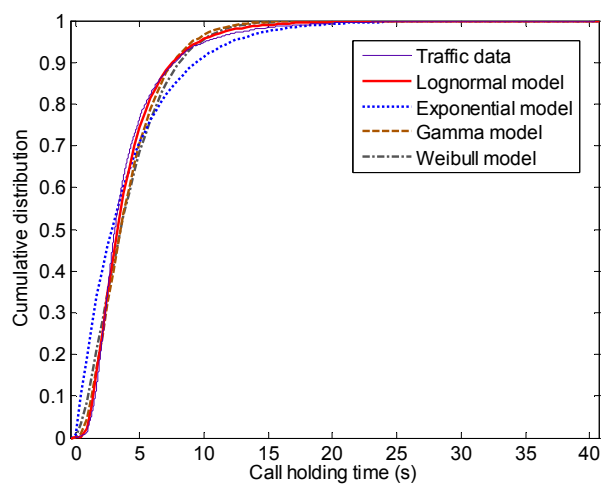
Input parameters

- Significance level σ: determines if the null hypothesis is wrongly rejected σ percent of times, if it is in fact true
  - default value $\sigma = 0.05$
- σ defines sensitivity of the test:
  - smaller σ implies larger critical value (larger tolerance)
- tail: specifies whether the K-S performs two sided test (default) or tests from one or other side of the candidate distribution

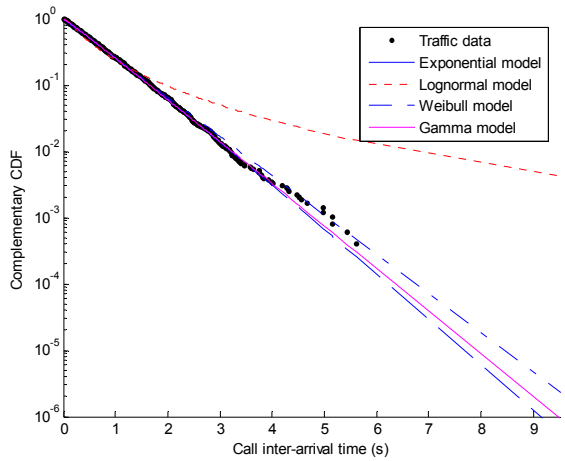
Output parameters

- Test statistic k is the maximum difference over all data points: $k = \max_{1 \le i \le N} \left| F(Y_i) - \frac{i}{N} \right|$
  where F is the CDF of the assumed distribution
- The null hypothesis is accepted if the value of the test statistic is smaller than the critical value
- p-value is probability level when the difference between distributions (test statistics) becomes significant:
  - if p-value $\leqslant \sigma$: test rejects the null hypothesis
- If test returns critical value = NaN, the decision to accept or reject null hypothesis is based only on p-value

Best-fitting distributions: CDF



Inter-arrival time: complementary CDF



K-S test: call inter-arrival times 2001

Significance level $\sigma = 0.1$

Distribution	Parameter	02.11.2001, 20:00–21:00	02.11.2001, 16:00–17:00	02.11.2001, 15:00–16:00	01.11.2001, 19:00–20:00	01.11.2001, 00:00–01:00
exponential	h	1	1	0	1	1
	p	0.0384	0.0001	0.5416	0.0122	0.0135
	k	0.0247	0.0369	0.0131	0.0277	0.0259
Weibull	h	0	1	0	0	1
	p	0.3036	0.0409	0.4994	0.1574	0.0837
	k	0.0171	0.0236	0.0136	0.0195	0.0206
gamma	h	0	1	0	1	1
	p	0.3833	0.0062	0.3916	0.0644	0.0953
	k	0.0159	0.0287	0.0148	0.0227	0.0202

Significance level σ	0.01	0.04	0.05	0.08	0.09	0.1
02.11.2001, 16:00–17:00: cv	0.0275	0.0237	0.0230	0.0215	0.0211	0.0207
01.11.2001, 00:00–01:00: cv	0.0267	0.0229	0.0223	0.0208	0.0204	0.0201

Traffic data

- Records of network events:
  - established, queued, and dropped calls in the Vancouver cell
- Traffic data span periods during:
  - 2001, 2002, 2003

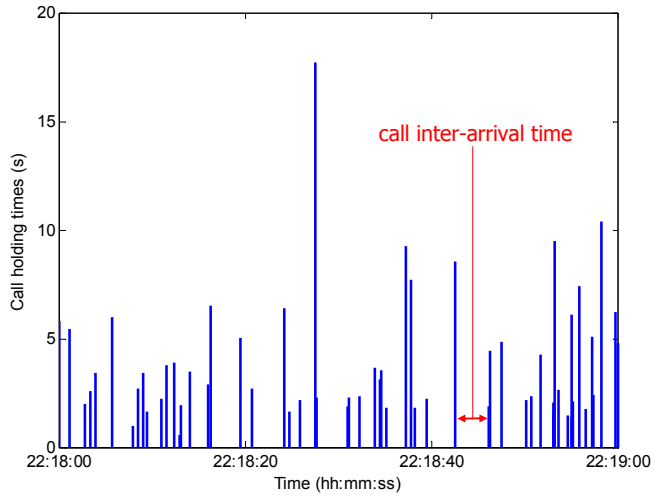
Trace (dataset)	Time span	No. of established calls
2001	November 1–2, 2001	110,348
2002	March 1–7, 2002	370,510
2003	March 24–30, 2003	387,340

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

Example: March 26, 2003



Statistical distributions

- Fourteen candidate distributions:
  - exponetial, 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
  - K-S test on potential candidates

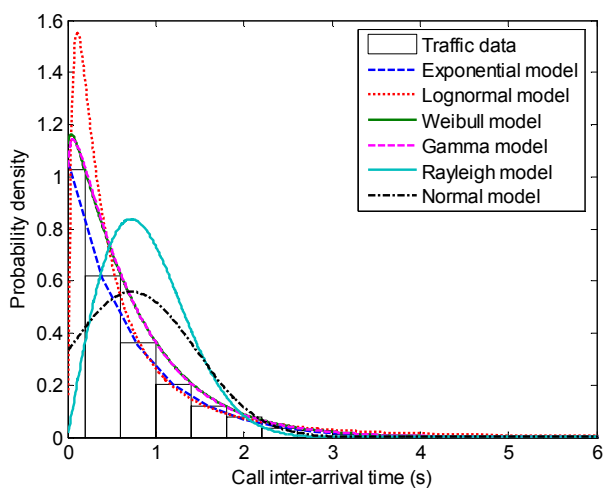
Maximum Likelihood Estimation (MLE)

- Introduced by R. A. Fisher in 1920s
- The most popular method for parameter estimation
- Goal: to find the distribution parameters that make the given distribution that follow the most closely underlying data set
- Conduct an experiment and obtain *N* independent observations
- $\theta_1, \theta_2, \ldots, \theta_k$ are *k* unknown constant parameters which

$$L(x_1, x_2, \ldots, x_N | \theta_1, \theta_2, \ldots, \theta_k) = L = \prod_{i=1}^{N} f(x_i; \theta_1, \theta_2, \ldots, \theta_k)$$

$$i = 1, 2, \ldots, N$$

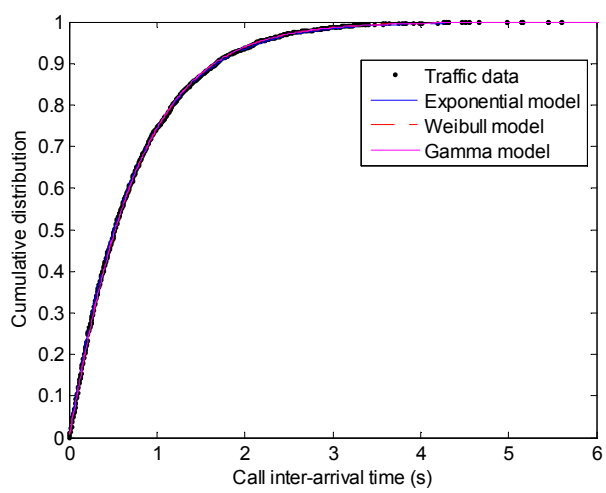
Call inter-arrival times: pdf candidates



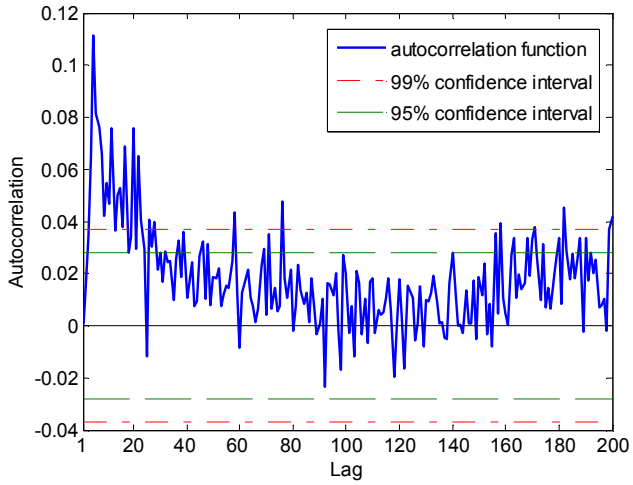
K-S test results: 2003

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	0.013	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

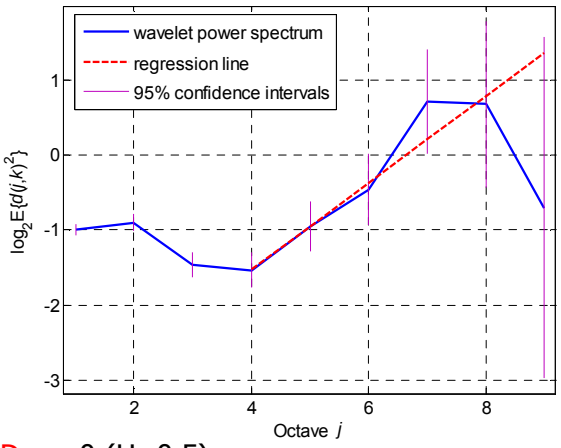
Call inter-arrival times, best-fitting distributions: cdf



Call inter-arrival time: autocorrelation



Logscale diagram, call inter-arrival times: 26.03.2003, 22:00–23:00



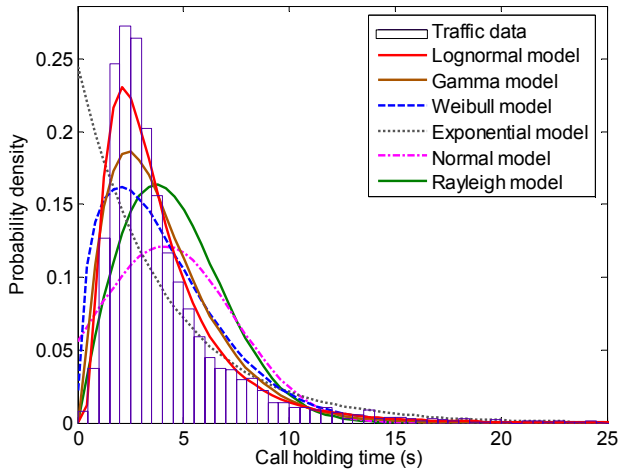
- LRD: $\alpha>0$ (H>0.5)
- similar logscale diagrams for other traces

Call inter-arrival times: estimates of H

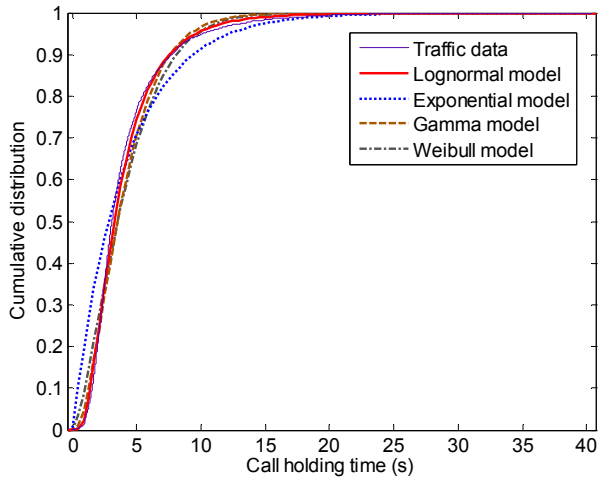
- Traces pass the test for time constancy of α: estimates of H are reliable

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 time: pdf candidates



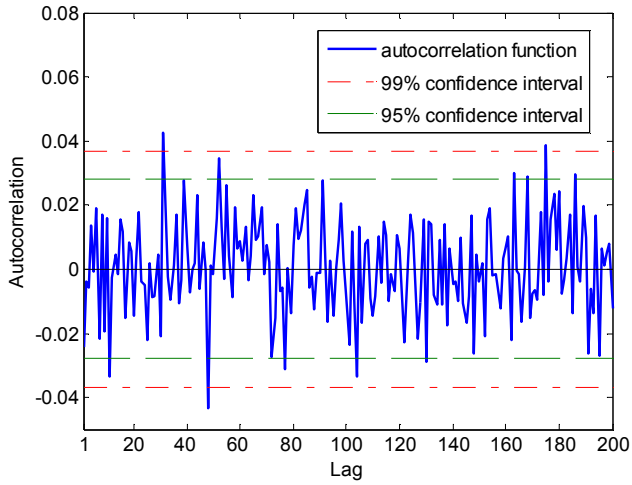
Best-fitting distributions: cdf



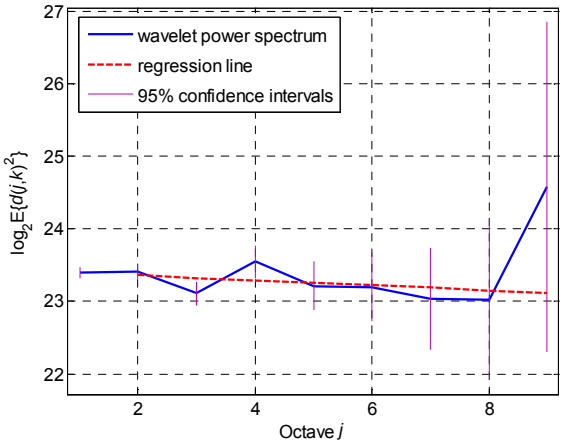
K-S test results: 2003

- No distribution passes the test when the entire trace is tested (significance levels = 0.1 and 0.01)
- Lognormal distribution passes test (significance level = 0.01) for:
  - 5-6 sub-traces from 15 randomly chosen 1,000-sample sub-traces
  - passes the test for almost all 500-sample sub-traces
- Test rejects null hypothesis when the sub-traces are compared with candidate distributions:
  - exponential
  - Weibull
  - gamma

Call holding time: autocorrelation



Logscale diagram, call holding times: 26.03.2003, 22:00–23:00



- independence: $\alpha \approx 0$ ($H \approx 0.5$)
- similar logscale diagrams for other traces

Call holding times: estimates of H

- All (except one) traces pass the test for constancy of α
- 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)

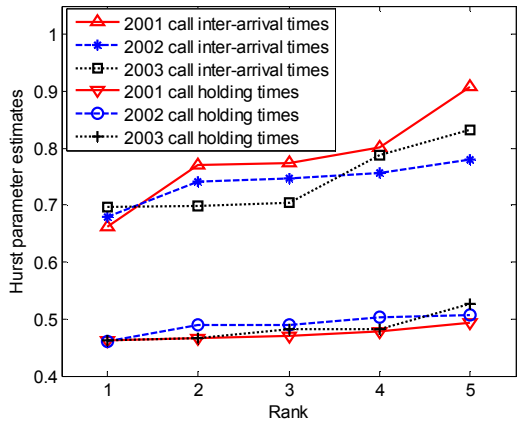
Distributions

Distribution	Expression	Remark
exponential	$f(x) = \frac{e^{-x/\mu}}{\mu}$	
Weibull	$f(x) = ba^{-b}x^{b-1}e^{-(x/a)^b} I_{(0,\infty)}(x)$	$I_{(0,\infty)}(x)$: incomplete beta function
gamma	$f(x) = \frac{x^{a-1}e^{-(x/b)}}{b^a \Gamma(a)}$	$\Gamma(a)$: gamma function
lognormal	$f(x) = \frac{e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}}{x\sigma\sqrt{2\pi}}$	

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

Estimates of H



- call inter-arrival times: H≈0.7–0.8
- call holding times: H≈0.5

Modeling and characterization of traffic: summary

- 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

Modeling and characterization of traffic: summary

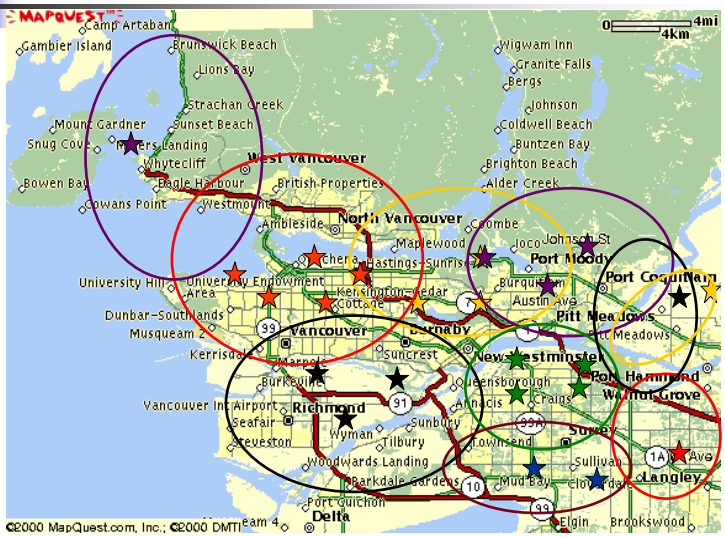
- Call inter-arrival times:
  - best fit: Weibull and gamma distributions
  - long-range dependent: $H \approx 0.7–0.8$
- Call holding times:
  - best fit: lognormal distribution
  - uncorrelated

Traffic prediction

- Trunked radio systems: traffic prediction based on user clusters
- Traffic data
- User clustering
- Traffic prediction

H. Chen and Lj. Trajković, "Trunked radio systems: traffic prediction based on user clusters," in *Proc. IEEE ISWCS 2004*, Mauritius, Sept. 2004, pp. 76–80.

E-Comm network coverage



Traffic data

- Collected data contain continuous data records from 92 days: March 1st 2003 – May 31st 2003
- Original database: ~6 GBytes, with 44,786,489 record rows
- 9 (out of original 26) fields are of interest for our analysis
- Data pre-processing:
  - cleaning the database
  - filtering the outliers
  - removing redundant records
  - extracting accurate user calling activity
- After the data cleaning and extraction, the resulting database contained ~20% of the original data

Traffic data: sample

Date	Time	Ms	Duration	Sys_id	Chl_id	Caller	Callee	C_type	C_state	Multi
2003-03-20	00:00:01	450	3730	8	4	6155	1801	0	0	0
2003-03-20	00:00:01	469	3730	6	7	6155	1801	0	0	0
2003-03-20	00:00:01	560	3730	3	7	6155	1801	0	0	0
2003-03-20	00:00:01	570	3730	2	7	6155	1801	0	0	0
2003-03-20	00:00:01	640	3730	1	7	6155	1801	0	0	0
2003-03-20	00:00:01	880	5260	9	6	13314	251	0	0	0
2003-03-20	00:00:01	910	5260	7	6	13314	251	0	0	0
2003-03-20	00:00:01	970	5260	6	8	13314	251	0	0	0
2003-03-20	00:00:01	980	2520	7	7	13911	418	0	0	0
2003-03-20	00:00:02	29	5270	4	2	13314	251	0	0	0
2003-03-20	00:00:02	109	5260	2	8	13314	251	0	0	0
2003-03-20	00:00:02	139	5270	1	8	13314	251	0	0	0
2003-03-20	00:00:02	9	2510	6	1	13911	418	0	0	0
2003-03-20	00:00:02	149	2510	2	9	13911	418	0	0	0
2003-03-20	00:00:05	289	3560	8	5	6011	2035	0	0	0
2003-03-20	00:00:05	309	3550	6	3	6011	2035	0	0	0
2003-03-20	00:00:05	389	3560	3	2	6011	2035	0	0	0
2003-03-20	00:00:05	449	3550	2	2	6011	2035	0	0	0
2003-03-20	00:00:05	480	3550	1	9	6011	2035	0	0	0
2003-03-20	00:00:05	550	3440	1	12	7614	945	0	0	0
2003-03-20	00:00:05	550	3440	2	3	7614	945	0	0	0
2003-03-20	00:00:05	949	9780	6	4	15840	418	0	0	0
2003-03-20	00:00:05	959	9780	7	2	15840	418	0	0	0
2003-03-20	00:00:06	679	3040	2	6	13931	471	0	0	0
2003-03-20	00:00:06	709	3040	1	2	13931	471	0	0	0
2003-03-20	00:00:06	130	9780	2	4	15840	418	0	0	0
2003-03-20	00:00:08	109	6640	9	2	13420	251	0	0	0
2003-03-20	00:00:08	179	6630	7	3	13420	251	0	0	0
2003-03-20	00:00:08	200	6640	6	5	13420	251	0	0	0
2003-03-20	00:00:08	270	6630	4	5	13420	251	0	0	0
2003-03-20	00:00:08	329	6640	1	4	13420	251	0	0	0
2003-03-20	00:00:08	340	6640	2	7	13420	251	0	0	0

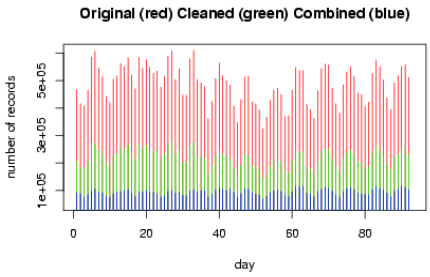
Traffic data: sample

- Traffic data after cleaning and extraction:

Date	Time	Ms	Duration	Caller	Callee	C_type	C_state	Multi	# Sys	System List
2003-03-20	00:00:01	450	3730	6155	1801	0	0	0	5	8,6,3,2,1
2003-03-20	00:00:01	980	2520	13911	418	0	0	0	3	7,6,2
2003-03-20	00:00:01	880	5260	13314	251	0	0	0	6	9,7,6,4,2,1
2003-03-20	00:00:05	550	3440	7614	945	0	0	0	2	1,2
2003-03-20	00:00:05	289	3560	6011	2035	0	0	0	5	8,6,3,2,1
2003-03-20	00:00:05	949	9780	15840	418	0	0	0	3	6,7,2
2003-03-20	00:00:06	810	2350	8022	817	0	0	0	1	1
2003-03-20	00:00:06	819	1590	13902	497	0	0	0	4	10,9,8,4
2003-03-20	00:00:06	440	3030	13931	471	0	0	0	5	10,9,4,2,1
2003-03-20	00:00:08	109	6640	13420	251	0	0	0	6	9,7,6,4,1,2

Data preparation

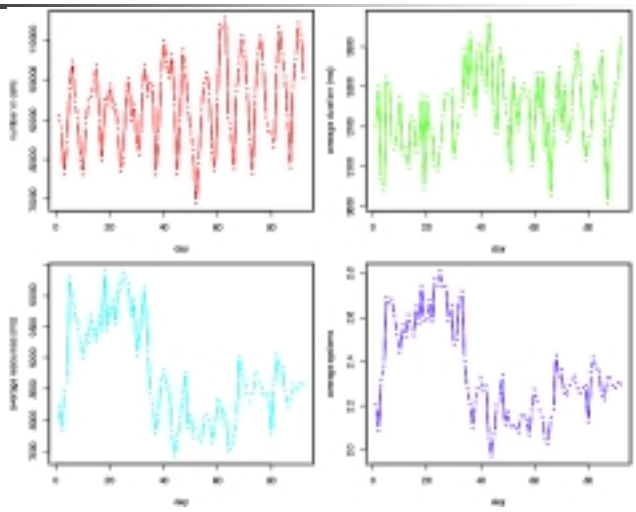
Date	Original	Cleaned	Combined
2003/03/01	466,862	204,357	91,143
2003/03/02	415,715	184,973	88,014
2003/03/03	406,072	182,311	76,310
2003/03/04	464,534	207,016	84,350
2003/03/05	585,561	264,226	97,714
2003/03/06	605,987	271,514	104,715
2003/03/07	546,230	247,902	94,511
2003/03/08	513,459	233,982	90,310
2003/03/09	442,662	201,146	79,815
2003/03/10	419,570	186,201	76,197
2003/03/11	504,981	225,604	88,857
2003/03/12	516,306	233,140	94,779
2003/03/13	561,253	255,840	95,662
2003/03/14	550,732	248,828	99,458
Total 92 Days	44,786,489	20,130,718	8,663,586
		44.95%	19.34%



Talk group data

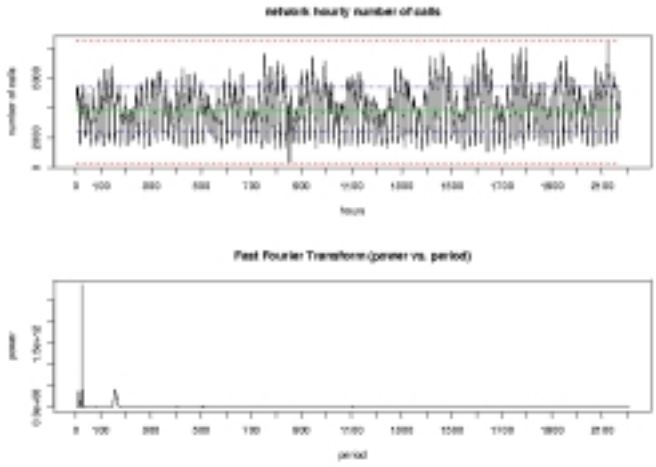
- The hourly number of calls is a common metric employed in telecommunication industry
- It is a footprint of users' calling behavior
- Time scales:
  - finer than an hour are too small to record the calling activity
  - larger than an hour are too coarse to capture user behavior patterns
- Traffic data were collected during 92 days (2,208 hours)
- The 2,208 hourly calls captured each talk group's calling behavior

Aggregate behavior

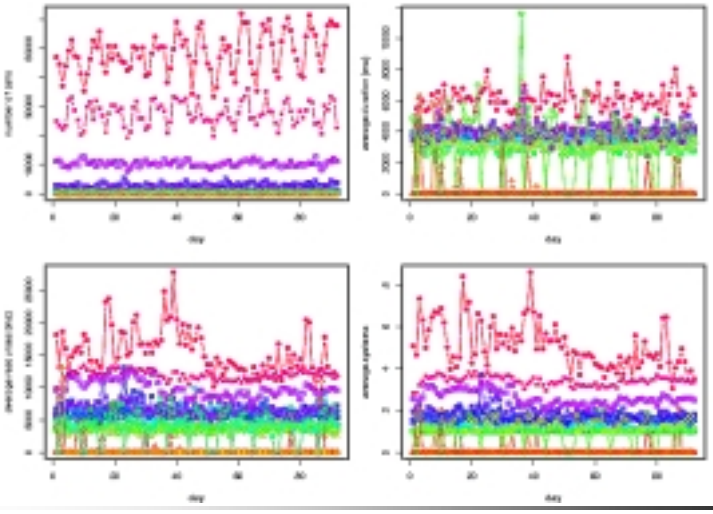


resources = number of calls × number of systems (cells)

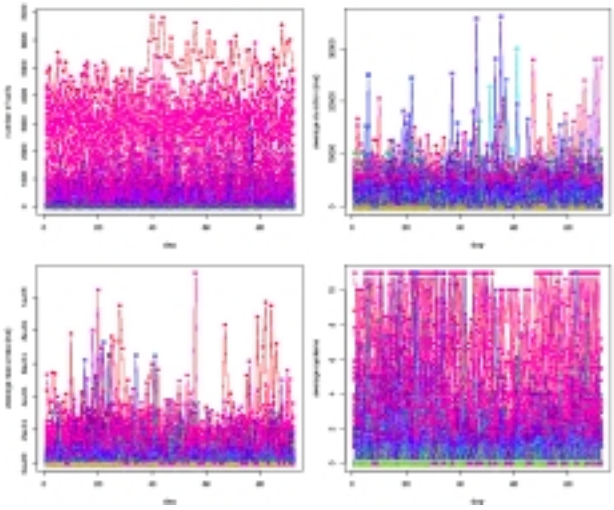
Periodicity



Agency behavior

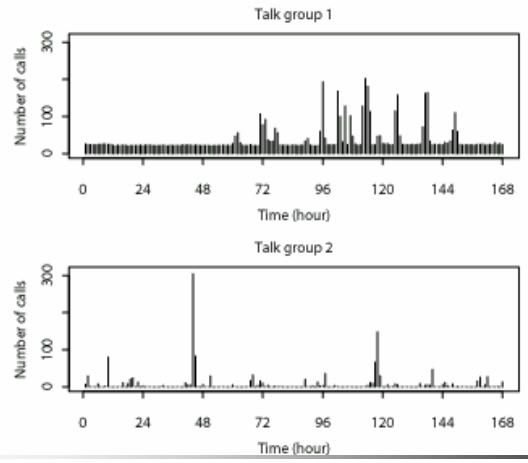


Talk groups behavior

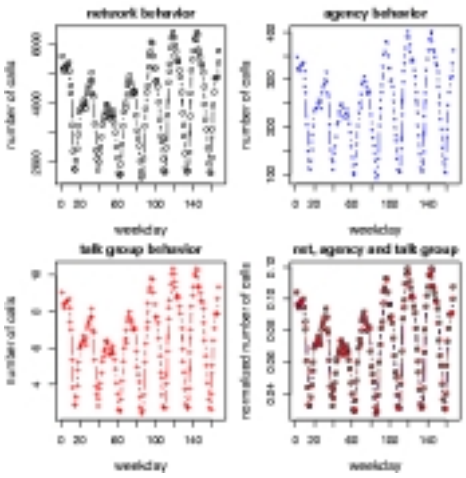


Calling patterns: talk groups 1 and 2

- Calling patterns are used for classification:



Data analysis: summary



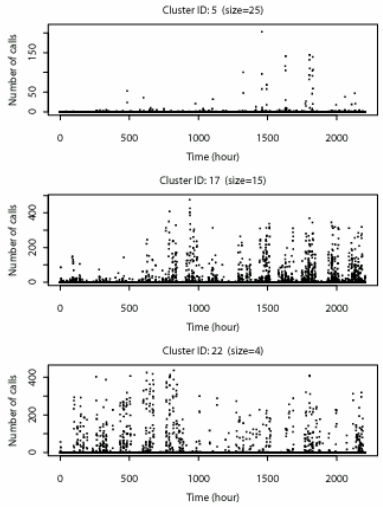
AutoClass: cluster size

- After 1,100 tries over 18 hours, AutoClass discovered 24 populated clusters of talk groups as the best classification, with the average cluster size of 25.7:

Cluster ID	1	2	3	4	5	6	7	8	9	10	11	12
Size	144	67	66	31	25	23	22	21	20	30	19	19

Cluster ID	13	14	15	16	17	18	19	20	21	22	23	24
Size	18	18	18	17	15	13	12	10	9	4	3	3

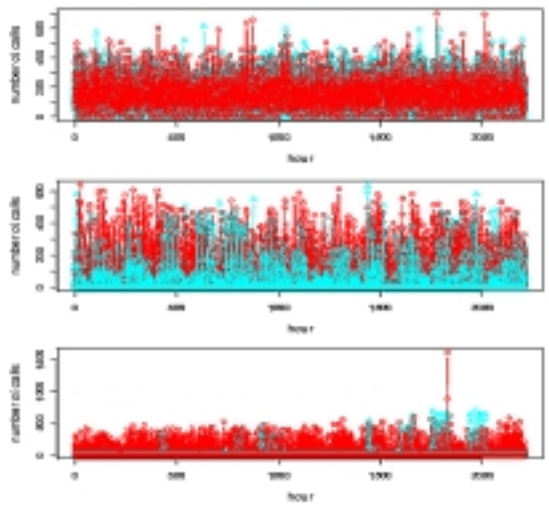
AutoClass: clusters 5, 17, and 22



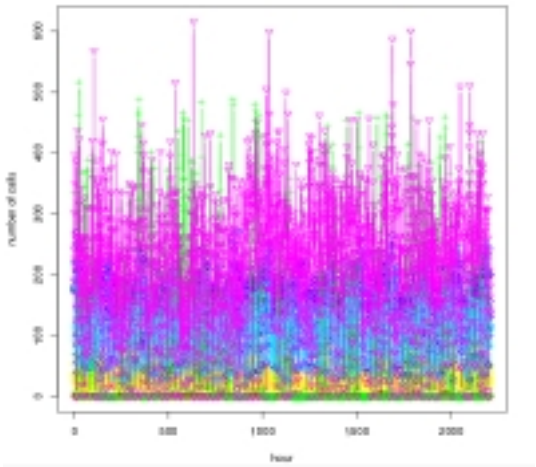
User clustering with K-means: k = 3

- First cluster (heavy network users):
  - 17 talk groups, contributing to 59% of the calls, with an average number of calls ranging from 94 to 208 per hour
  - They are dispatch groups that assign and schedule other talk groups for specific tasks
- Second cluster (average network users):
  - 31 talk groups, contributing to 26% of the calls
- Third cluster (least frequent network users):
  - 569 talk groups, contributing to only 15% of the calls
  - They represent over 90% of all talk groups

User clusters with K-means: k = 3



User clusters with K-means: k = 6



K-means: cluster properties for k = 3

Cluster size	Range of min. # of calls	Range of max. # of calls	Range of avg. # of calls	Total # of calls	Number of calls (%)
17	0 – 6	352 – 700	94 – 208	5,091,695	59
31	0 – 3	135 – 641	17 – 66	2,261,055	26
569	0	1 – 1613	0 – 16	1,310,836	15

Prediction based on aggregate traffic

- The aggregate network traffic consists of all network users' traffic
- The R system was used to identify, estimate, and verify the SARIMA model for the aggregate users' traffic
- Both 24-hour (one day) and 168-hour (one week) intervals were selected as seasonal period parameters
- Based on m past traffic data samples, we forecast the future n traffic data samples
- The prediction quality was measured using the normalized mean square error nmse:

$$nmse(a,b) = \sum_{i=m+1}^{m+n} \frac{(a_i - b_i)^2}{a_i^2},$$

- where: a_i is the observed and b_i is the predicted data

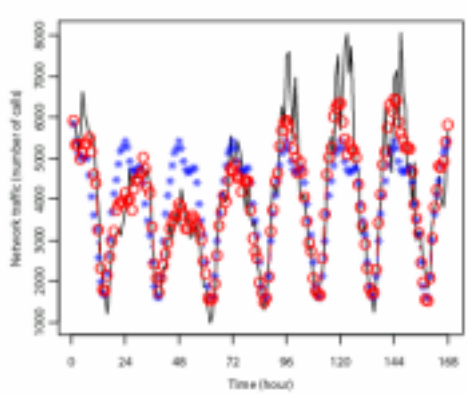
Prediction based on aggregate traffic

p	d	q	P	D	Q	S	m	n	nmse
2	0	1	0	1	1	24	1,920	24	0.1941
3	0	1	0	1	1	24	1,920	24	0.1907
2	0	1	0	1	1	24	1,680	168	0.4079
3	0	1	0	1	1	24	1,680	168	0.4081
2	0	1	0	1	1	168	1,920	24	0.0969
3	0	1	0	1	1	168	1,920	24	0.1012
2	0	1	0	1	1	168	1,680	168	0.1745
3	0	1	0	1	1	168	1,680	168	0.1748

Prediction based on aggregate traffic

- Two groups of models, with 24-hour and 168-hour seasonal periods:
  - ARIMA (2, 0, 1) x $(0, 1, 1)_{24}$
  - ARIMA (3, 0, 1) x $(0, 1, 1)_{168}$
- The four models with a 168-hour seasonal period provided better prediction than the four 24-hour period based models, particularly when predicting long term traffic data
- A drawback of the 168-hour based models is the large computational cost:
  - They require 200 $\times$ the CPU time required by the 24-hour based models

Prediction of 168 hours of traffic based on 1,680 past hours



Comparison of the 24-hour and the 168-hour models

- Solid line: observation
- o: prediction of 168-hour seasonal model
- *: prediction of 24-hour seasonal model

Prediction with user clustering

- Raw network log data collected over 92 days: March 1st 2003 – May 31st 2003
- Footprint of network usage for talk groups: the hourly number of calls
- AutoClass and the K-means algorithm were used to classify network talk groups into clusters
- The behavior of each user cluster was predicted using Seasonal Autoregressive Integrated Moving Average (SARIMA)
- We used aggregation to predict the overall network behavior

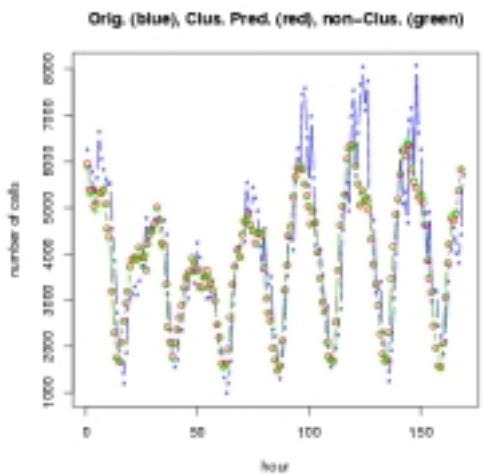
Traffic prediction based on user clusters

- The developed aggregate users based prediction assumes that the adopted model is static: the number of network users and their behavior pattern are constant in time
- This assumption does not hold when planning further network expansions and cannot be used to forecast network traffic
- We employed a user clusters based prediction approach to predict the network traffic by accumulating the prediction results from user clusters
- In large networks with many users, it is impractical to predict individual users' traffic and then aggregate the predicted data
- With user clusters, traffic prediction is reduced to predicting and aggregating users' traffic from few clusters

Prediction of 48 hours of traffic based on 1,680 past hours

Comparisons: model $(2,0,1)x(0,1,1)_{24}$
* observation
* prediction without clustering
o prediction with clustering
+ optimized prediction with clustering

Prediction of 168 hours of traffic based on 1,680 past hours



Comparisons: model $(1,0,1)x(0,1,1)_{168}$
* observation
* prediction without clustering
o prediction with clustering

Prediction results with user clusters

Cluster	(p,d,q)	(P,D,Q)	S	m	n	nmse
1	(2,0,1)	(0,1,1)	24	1,680	48	1.1954
2	(2,0,1)	(0,1,1)	24	1,680	48	2.4519
3	(2,0,1)	(0,1,1)	24	1,680	48	0.3701
*	(2,0,1)	(0,1,1)	24	1,680	48	0.6298
A	(2,0,1)	(0,1,1)	24	1,680	48	0.6256
B	(2,0,1)	(0,1,1)	24	1,680	48	0.4231

Traffic prediction with user clusters

- Rows 1, 2, and 3: traffic prediction results for user clusters 1, 2, and 3
- Row *: the aggregate user traffic prediction obtained without clustering the users
- Row A: the weighted aggregate prediction of network traffic based on the three user clusters
- Row B: the optimized weighted aggregate prediction
- nmse > 1.0 for clusters 1 and 2 implies that the prediction results are worse than the prediction based on the mean value of past data
- Row B: a better prediction, achieved when the mean value prediction was used for clusters 1 and 2
- Row A: even the non-optimized cluster-based prediction is not worse than the prediction using aggregate traffic

Traffic prediction with user clusters: more examples: (2,0,1) x (0,1,1)

Cluster	S	m	n	nmse
1	168	1,920	24	0.2241
2	168	1,920	24	0.3818
3	168	1,920	24	0.1163
*	168	1,920	24	0.0969
A	168	1,920	24	0.1175
1	24	1,920	24	0.2508
2	24	1,920	24	0.2697
3	24	1,920	24	0.3020
*	24	1,920	24	0.1941
A	24	2,920	24	0.2052
1	24	1,680	168	0.5477
2	24	1,680	168	0.6883
3	24	1,680	168	0.2852
*	24	1,680	168	0.4079
A	24	1,680	168	0.4093

Prediction results with user clusters

- For each group, rows 1, 2, and 3: traffic prediction results for user clusters 1, 2, and 3
- Row *: the aggregate user traffic prediction obtained without clustering the users
- Row A: the aggregate prediction of network traffic based on the three user clusters
- The performance of the clusters based prediction (nmse: 0.1175) is comparable to the best prediction based on aggregate traffic (nmse: 0.0969)
- Prediction of traffic in networks with a variable number of users is possible, as long as the new user groups could be classified into the existing user clusters

Traffic prediction: summary

- We analyzed traffic data collected from an operational trunked radio network
- We used the K-means algorithm and AutoClass to classify network users into user clusters
- We predicted network traffic using the SARIMA model based on aggregate user traffic and based on three user clusters that produced better prediction results
- The user clusters based prediction approach is also applicable to networks with variable number of users where prediction based on aggregate traffic could not be applied
- Our approach may enable network operators to predict network traffic and may provide guidance for network expansion

Road map

- Introduction
- Networks and traffic data:
  - data collection
  - statistical analysis
  - traffic prediction
- Two case studies:
  - satellite network: ChinaSat
  - wireless network: E-Comm
- Conclusions and references

Conclusions

- We analyzed traffic data collected from two deployed networks:
  - ChinaSat
  - E-Comm
- ChinaSat, system:
  - Web applications and TCP protocol dominate the collected traffic
  - packet size distribution is bimodal and most bytes are transferred in large packets
  - few Web servers account for majority of data traffic
  - the frequency-rank relation of client connections matches the discrete lognormal distribution

Conclusions

- ChinaSat, traffic:
  - self-similarity (Hurst parameter) was estimated using various estimators
- ChinaSat, TCP model:
  - Weibull: inter-arrival times of TCP connections
  - Lognormal: downloaded bytes per TCP connection
- ChinaSat, traffic prediction with ARIMA model:
  - performs better for the uploaded traffic
  - not suitable for the downloaded traffic

Conclusions

- E-Comm, users:
  - Clustering algorithms (K-means and AutoClass) were employed to classify network users into user clusters
- E-Comm, traffic prediction:
  - Network traffic was predicted using the SARIMA model based on aggregate user traffic and based on three user clusters
  - The user clusters based prediction approach is applicable to networks with variable number of users

References: self-similarity

- P. Abry and D. Veitch, "Wavelet analysis of long-range dependence traffic," *IEEE Transactions on Information Theory*, vol. 44, no. 1, pp. 2–15, Jan. 1998.
- P. Abry, P. Flandrin, M. S. Taqqu, and D. Veitch, "Wavelets for the analysis, estimation, and synthesis of scaling data," in *Self-similar Network Traffic and Performance Evaluation*, K. Park and W. Willinger, Eds. New York: Wiley, 2000, pp. 39–88.
- P. Barford, A. Bestavros, A. Bradley, and M. Crovella, "Changes in Web client access patterns: characteristics and caching implications in world wide web," *World Wide Web*, Special Issue on Characterization and Performance Evaluation, vol. 2, pp. 15–28, 1999.
- Z. Bi, C. Faloutsos, and F. Korn, "The 'DGX' distribution for mining massive, skewed data," in *Proc. of ACM SIGCOMM Internet Measurement Workshop*, San Francisco, CA, Aug. 2001, pp. 17–26.
- M. E. Crovella and A. Bestavros, "Self-similarity in world wide web traffic: evidence and possible causes," *IEEE/ACM Transactions on Networking*, vol. 5, no. 6, pp. 835–846, Dec. 1997.

References: self-similarity

- A. Feldmann, "Characteristics of TCP connection arrivals," in *Self-similar Network Traffic and Performance Evaluation*, K. Park and W. Willinger, Eds., New York: Wiley, 2000, pp. 367–399.
- W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of ethernet traffic (extended version)," *IEEE/ACM Transactions on Networking*, vol. 2, no. 1, pp. 1–15, Feb. 1994.
- T. Karagiannis, M. Faloutsos, and R. H. Riedi, "Long-range dependence: now you see it, now you don't!," in *Proc. GLOBECOM '02*, Taipei, Taiwan, Nov. 2002, pp. 2165–2169.
- M. S. Taqqu and V. Teverovsky, "On estimating the intensity of long-range dependence in finite and infinite variance time series," in *A Practical Guide to Heavy Tails: Statistical Techniques and Applications.* Boston, MA: Birkhauser, 1998, pp. 177–217.

References: traffic analysis

- Y. W. Chen, "Traffic behavior analysis and modeling sub-networks," *International Journal of Network Management*, John Wiley & Sons, vol. 12, pp. 323–330, 2002.
- Y. Fang and I. Chlamtac, "Teletraffic analysis and mobility modeling of PCS networks," *IEEE Trans. on Communications*, vol. 47, no. 7, pp. 1062–1072, July 1999.
- N. K. Groschwitz and G. C. Polyzos, "A time series model of long-term NSFNET backbone traffic," in *Proc. IEEE International Conference on Communications* (ICC'94), New Orleans, LA, May 1994, vol. 3, pp. 1400–1404.
- D. Papagiannaki, N. Taft, Z.-L. Zhang, and C. Diot, "Long-term forecasting of Internet backbone traffic: observations and initial models," in *Proc. IEEE INFOCOM 2003,* San Francisco, CA, April 2003, pp. 1178–1188.
- D. Tang and M. Baker, "Analysis of a metropolitan-area wireless network," in *Proc. MOBICOM 1999*, Seattle, WA, USA, Aug. 1999, pp. 13–23.

References: traffic analysis

- F. Barceló and J. I. Sȧnchez, "Probability distribution of the inter-arrival time to cellular telephony channels," in *Proc. of the 49th Vehicular Technology Conference*, May 1999, vol. 1, pp. 762–766.
- F. Barcelo and J. Jordan, "Channel holding time distribution in public telephony systems (PAMR and PCS)," *IEEE Trans. Vehicular Technology*, vol. 49, no. 5, pp. 1615–1625, Sept. 2000.
- R. B. D'Agostino and M. A. Stephens, Eds., *Goodness-of-Fit Techniques*. New York: Marcel Dekker, 1986. pp. 63–93, pp. 97–145, pp. 421–457.

References: time series and clustering analysis

- G. Box and G. Jenkins, *Time Series Analysis: Forecasting and Control*, 2nd edition. San Francisco, CA: Holden-Day, 1976, pp. 208–329.
- P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, 2nd Edition. New York: Springer-Verlag, 2002.
- N. H. Chan, *Time Series: Applications to Finance*. New York: Wiley-Interscience, 2002.
- P. Cheeseman and J. Stutz, "Bayesian classification (AutoClass): theory and results," in *Advances in Knowledge Discovery and Data Mining*, U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, Eds., AAAI Press/MIT Press, 1996.
- J. W. Han and M. Kamber, *Data Mining: Concepts And Techniques*. San Francisco: Morgan Kaufmann, 2001.
- T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer, 2001.
- L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: John Wiley & Sons, 1990.

References: tools

- The R Project for Statistical Computing [Online]. Available: http://www.r-project.org.
- *OPNET documentation V.9.0.A*, OPNET Technologies, Inc., Bethesda, MD, 2001.
- OPNET [Online]. Available: http://www.opnet.com.
- WarnSim [Online]. Available: http://www.vannet.ca/warnsim.

References: case studies

- ChinaSat [Online]. Available: http://www.chinanex.com/company/chinasat.htm.
- DirecPC (currently DirecWay) [Online]. Available: http://www.direcway.com.
- E-Comm, Emergency Communications for Southwest British Columbia [Online]. Available: http://www.ecomm.bc.ca.
- EDACS Explained [Online]. Available: http://www.trunkedradio.net/trunked/edacs/EDACS_Whitepaper.pdf.
- EDACS Trunking Information. [Online]. Available: http://www.radioreference.com.
- In Vancouver! Vancouver Travel Guide. [Online]. Available: http://www.vancouver-bc.com/maps-html.

References: downloads

http://www.ensc.sfu.ca/~ljilja/publications_date.html

- N. Cackov, J. Song, B. Vujičić, S. Vujičić, and Lj. Trajković, "Simulation of a public safety wireless networks: a case study," *Simulation*, SCS, to appear.
- 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.
- J. Song and Lj. Trajković, "Modeling and performance analysis of public safety wireless networks," in *Proc. IEEE IPCCC*, Phoenix, AZ, Apr. 2005, pp. 567–572.
- H. Chen and Lj. Trajković, "Trunked radio systems: traffic prediction based on user clusters," in *Proc. IEEE ISWCS 2004*, Mauritius, Sept. 2004, pp. 76–80.
- D. Sharp, N. Cackov, N. Lasković, Q. Shao, and Lj. Trajković, "Analysis of public safety traffic on trunked land mobile radio systems," *IEEE J. Select. Areas Commun.*, vol. 22, no. 7, pp. 1197–1205, Sept. 2004.
- Q. Shao and Lj. Trajković, "Measurement and analysis of traffic in a hybrid satellite-terrestrial network," in *Proc. SPECTS 2004*, San Jose, CA, July 2004, pp. 329–336.
- N. Cackov, B. Vujičić, S. Vujičić, and Lj. Trajković, "Using network activity data to model the utilization of a trunked radio system," in *Proc. SPECTS*, San Jose, CA, July 2004, pp. 517–524.