MODELING AND CHARACTERIZATION OF TRAFFIC IN PUBLIC SAFETY WIRELESS NETWORKS

Božidar Vujičić, Nikola Cackov, Svetlana Vujičić, and Ljiljana Trajković
{bvujicic, ncackov, svujicic, ljilja}@cs.sfu.ca

Communication Networks Laboratory
http://www.ensc.sfu.ca/cnl
Simon Fraser University
Vancouver, BC

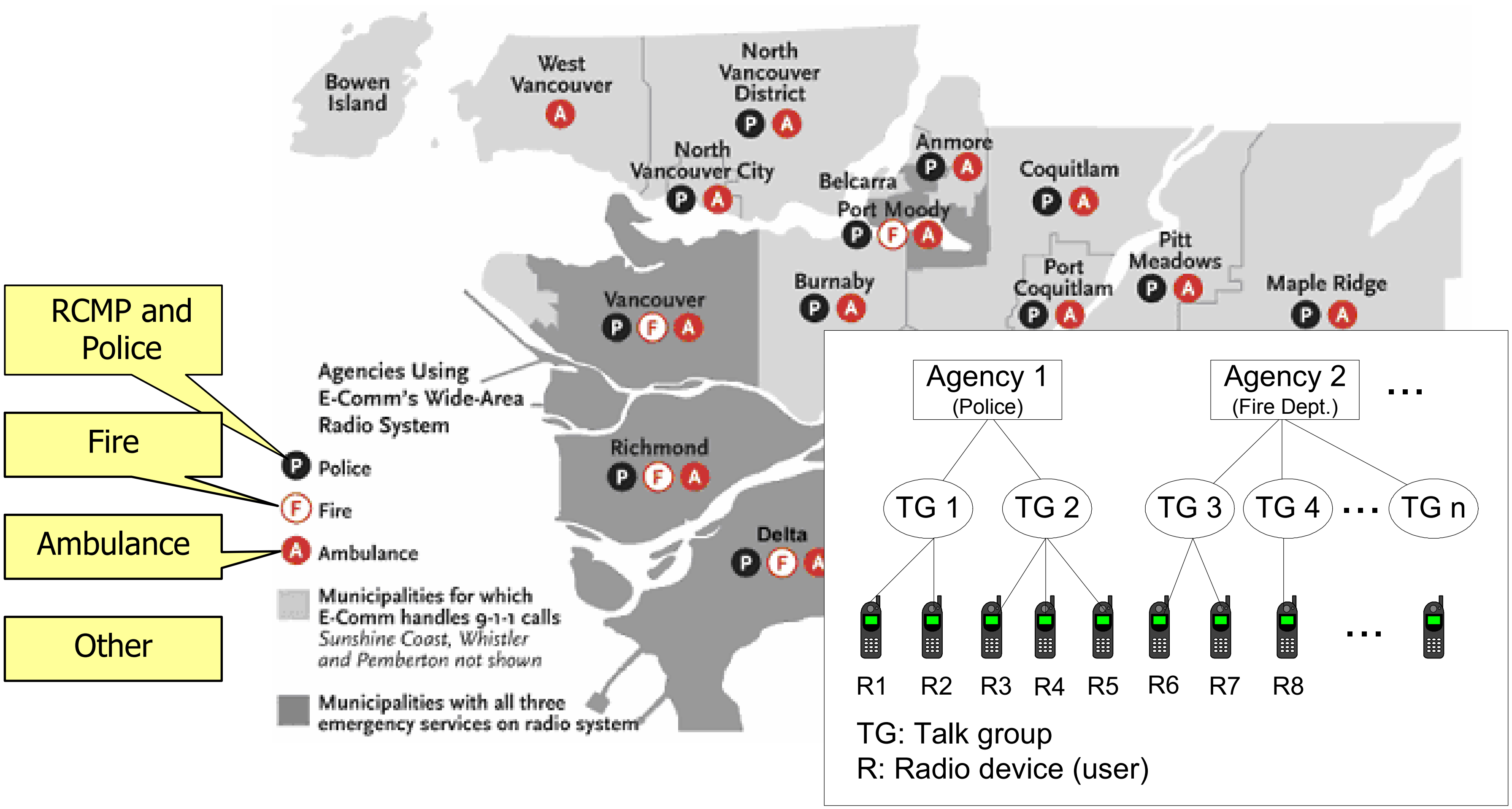
Roadmap

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

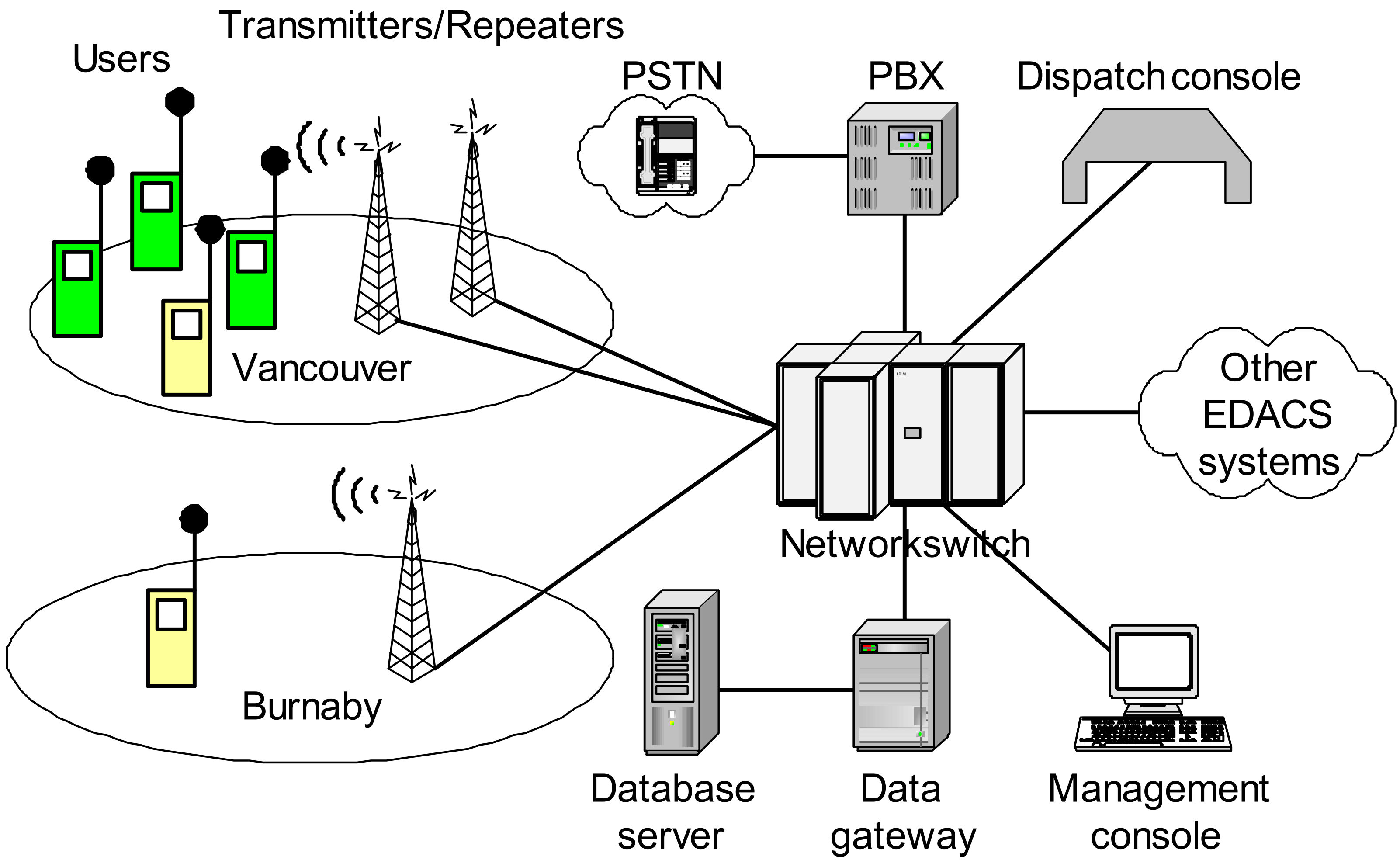
Roadmap

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

E-Comm network: coverage and user agencies



E-Comm network architecture



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

Previous work

- Simulation:
  - OPNET
  - WarnSim
- Traffic prediction based on user clusters
  - Seasonal ARIMA model
- Statistical analysis of traffic
  - three busy hours in 2001

[1] 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.

[2] 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.

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

[4] 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.

Roadmap

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

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

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

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:

$$\log_2 \mathrm{E}\{d(j,k)^2\} = \alpha \times j + c$$

- 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 E{d(j,k)^2} is a sample mean:

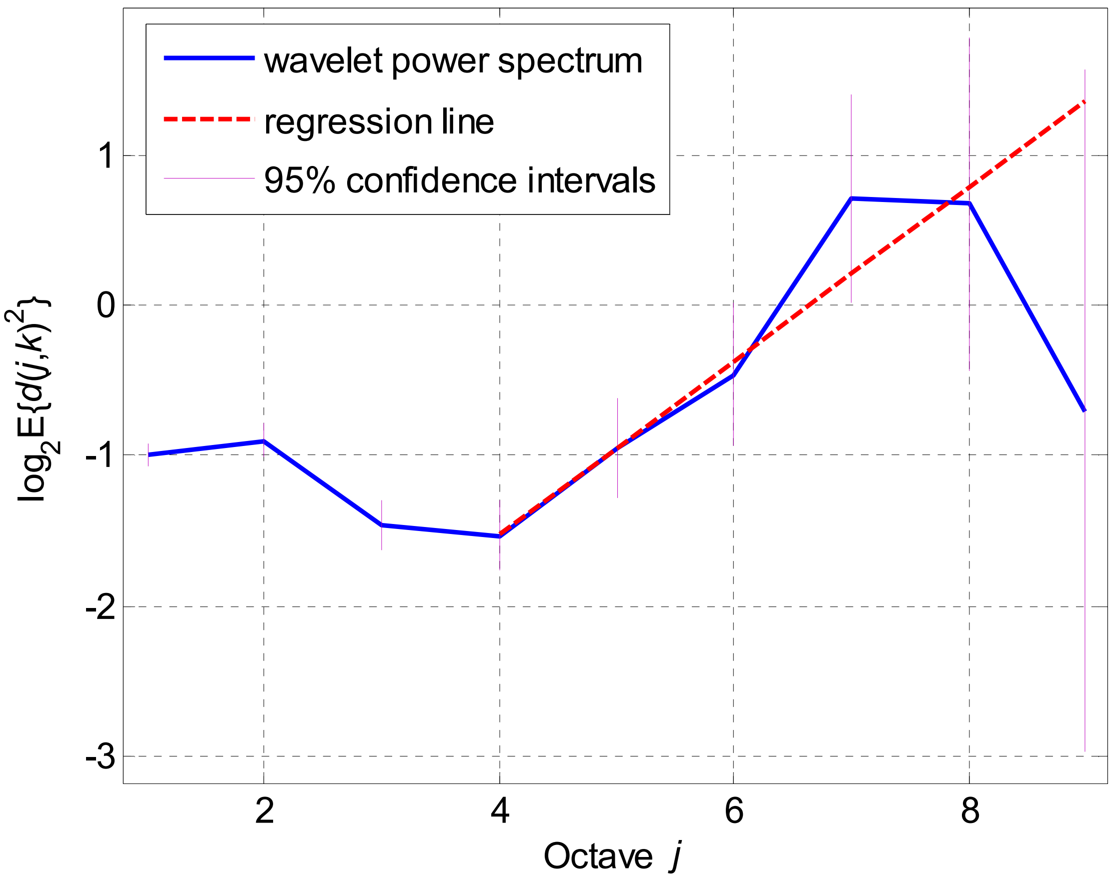
$$\mathrm{E}\{d(j,k)^2\} = \frac{1}{n_j}\sum_{k=1}^{n_j} d(j,k)^2$$

  - n_j: number of wavelet coefficients at octave j

Estimation of α and H

- Logscale diagram: plot of $\log_2 E\{d(j,k)^2\}$ 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, $m \in \{3, 4, 5, 6, 7, 8, 10\}$

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

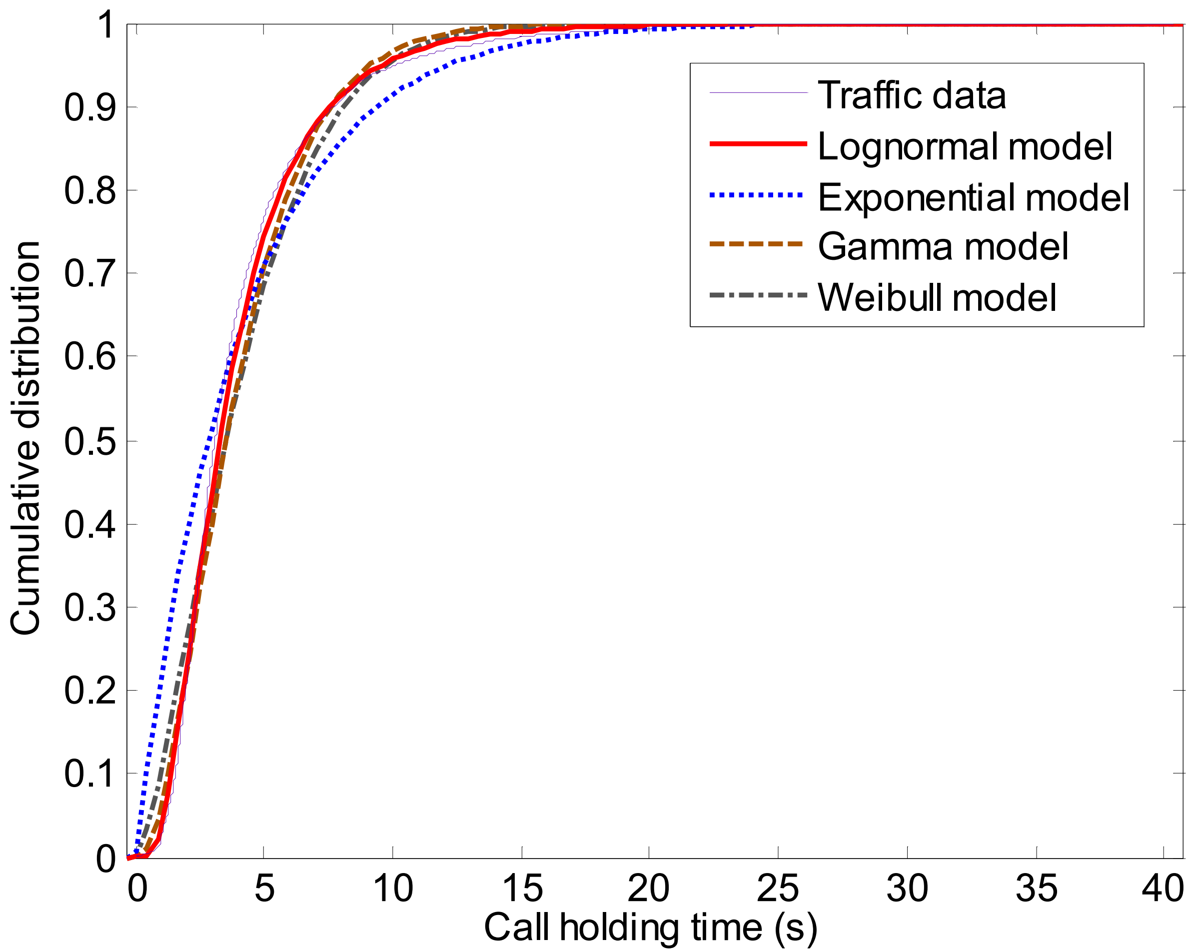
Input parameters

- 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

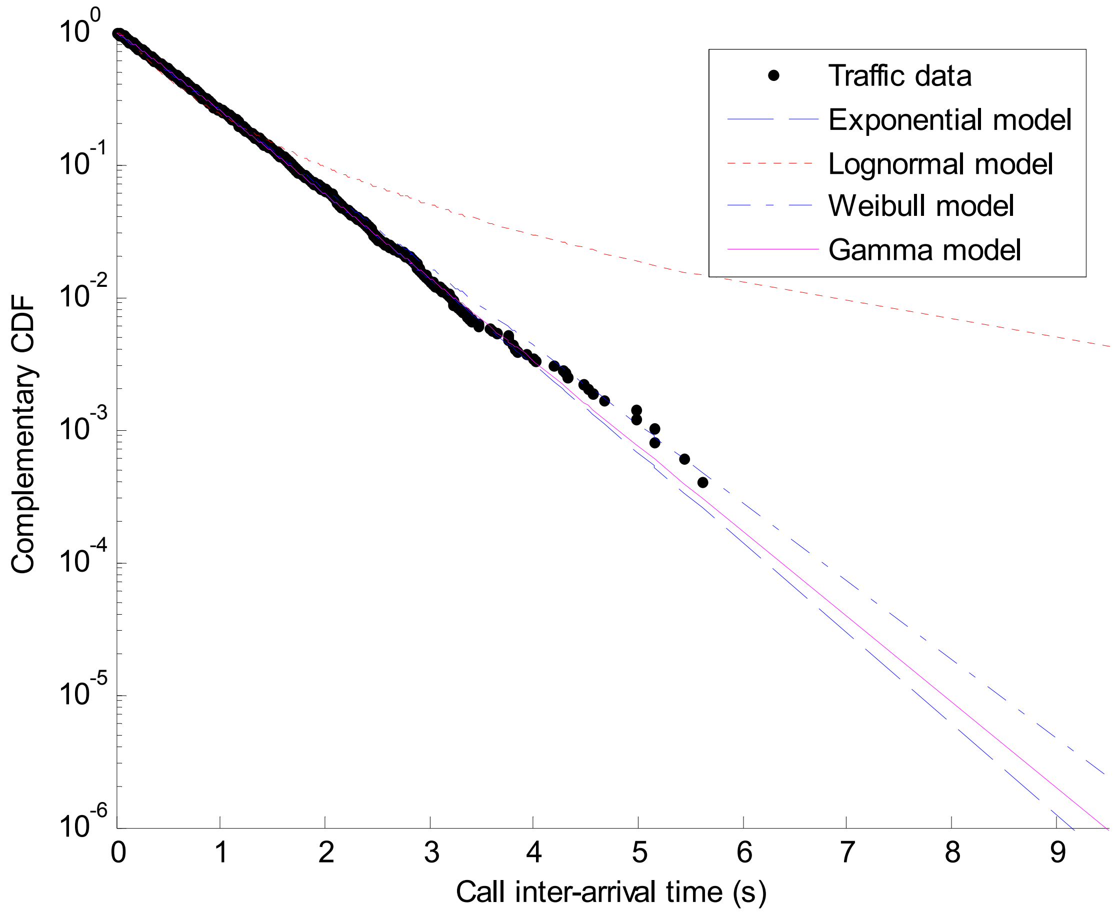
Output parameters

- 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

Roadmap

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

Traffic data

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

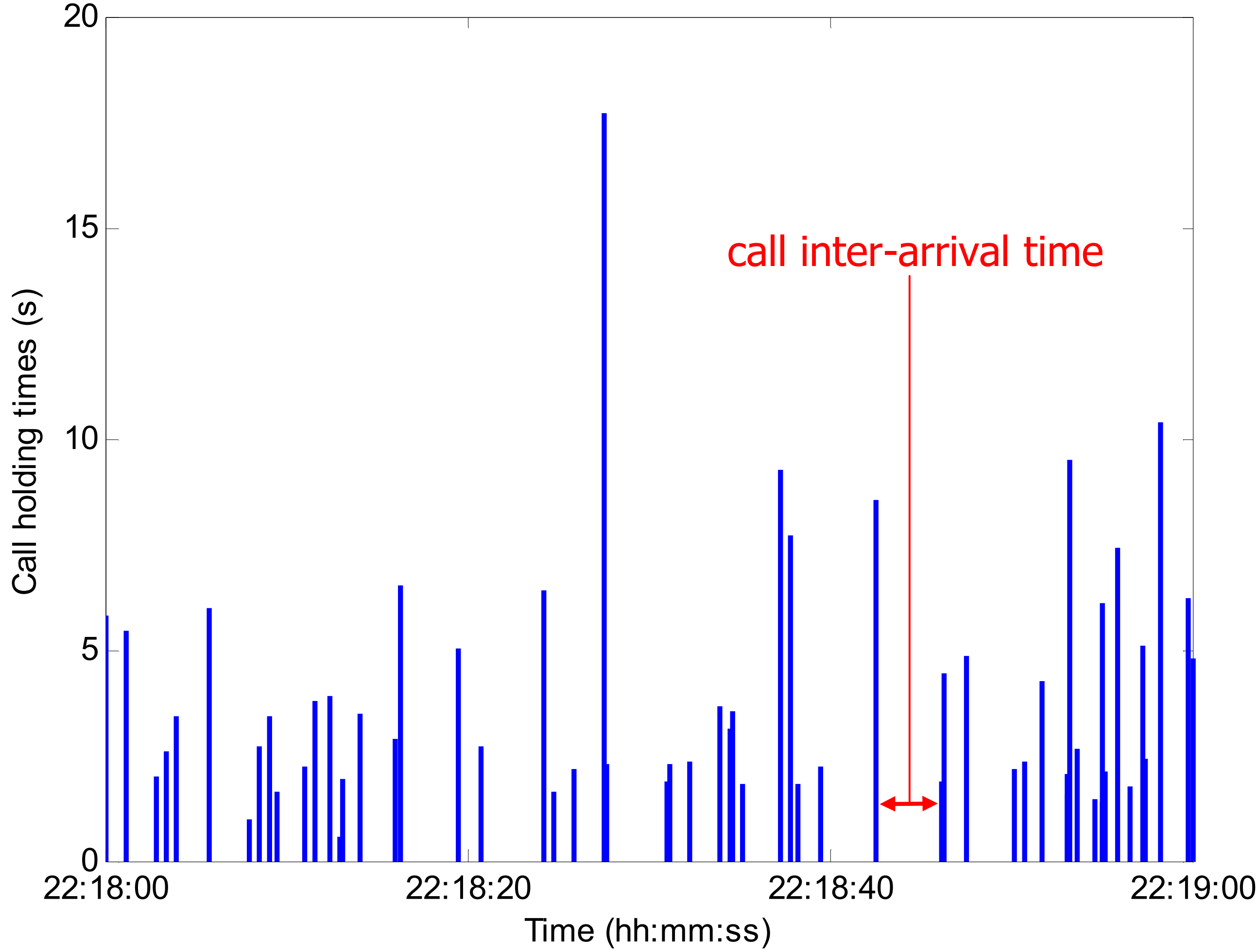
Trace (dataset)	Time span	No. of established calls
2001	November 1–2, 2001	110,348
2002	March 1–7, 2002	370,510
2003	March 24–30, 2003	387,340

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



Roadmap

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

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

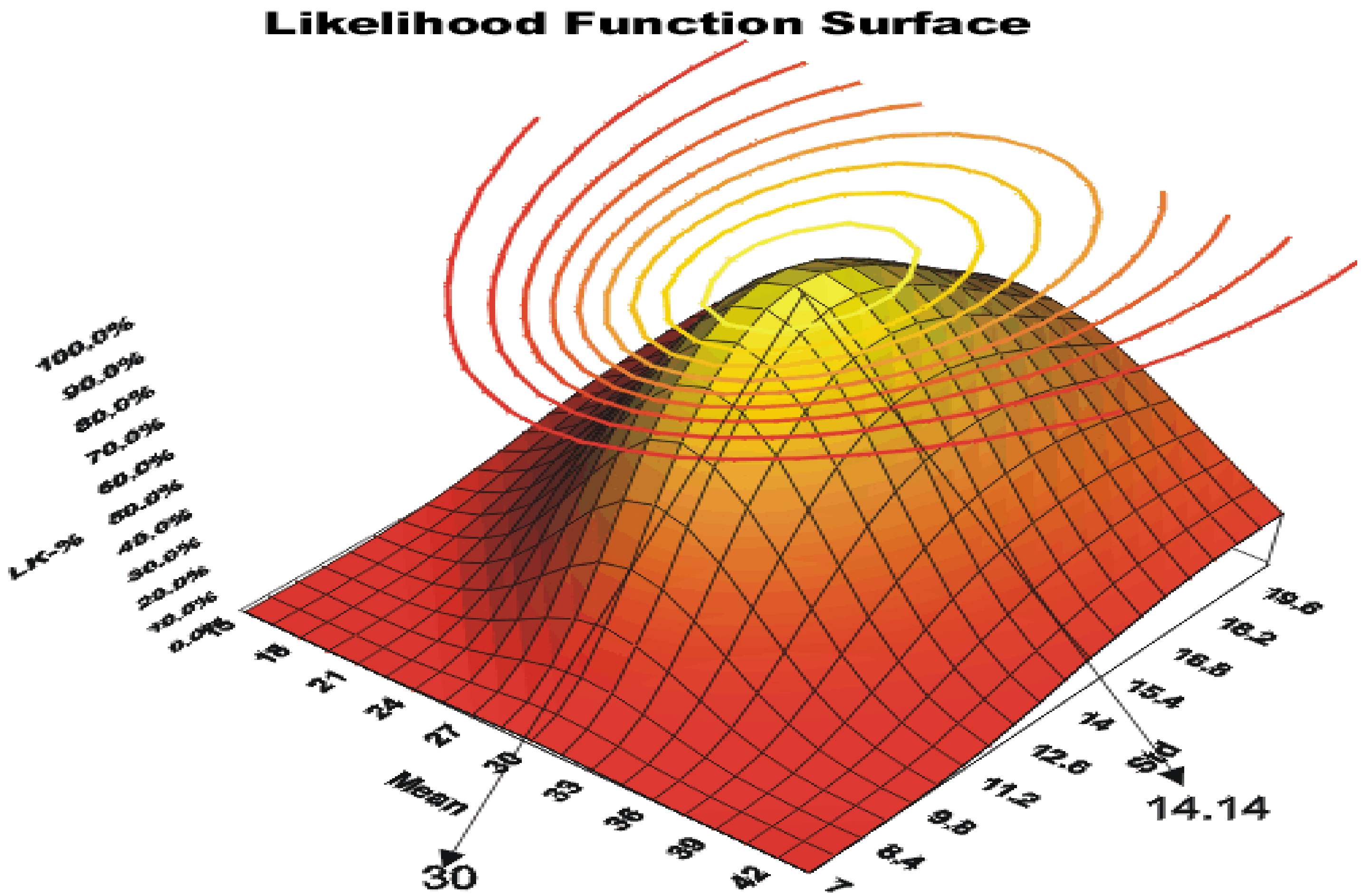
Maximum Likelihood Estimation (MLE)

- 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

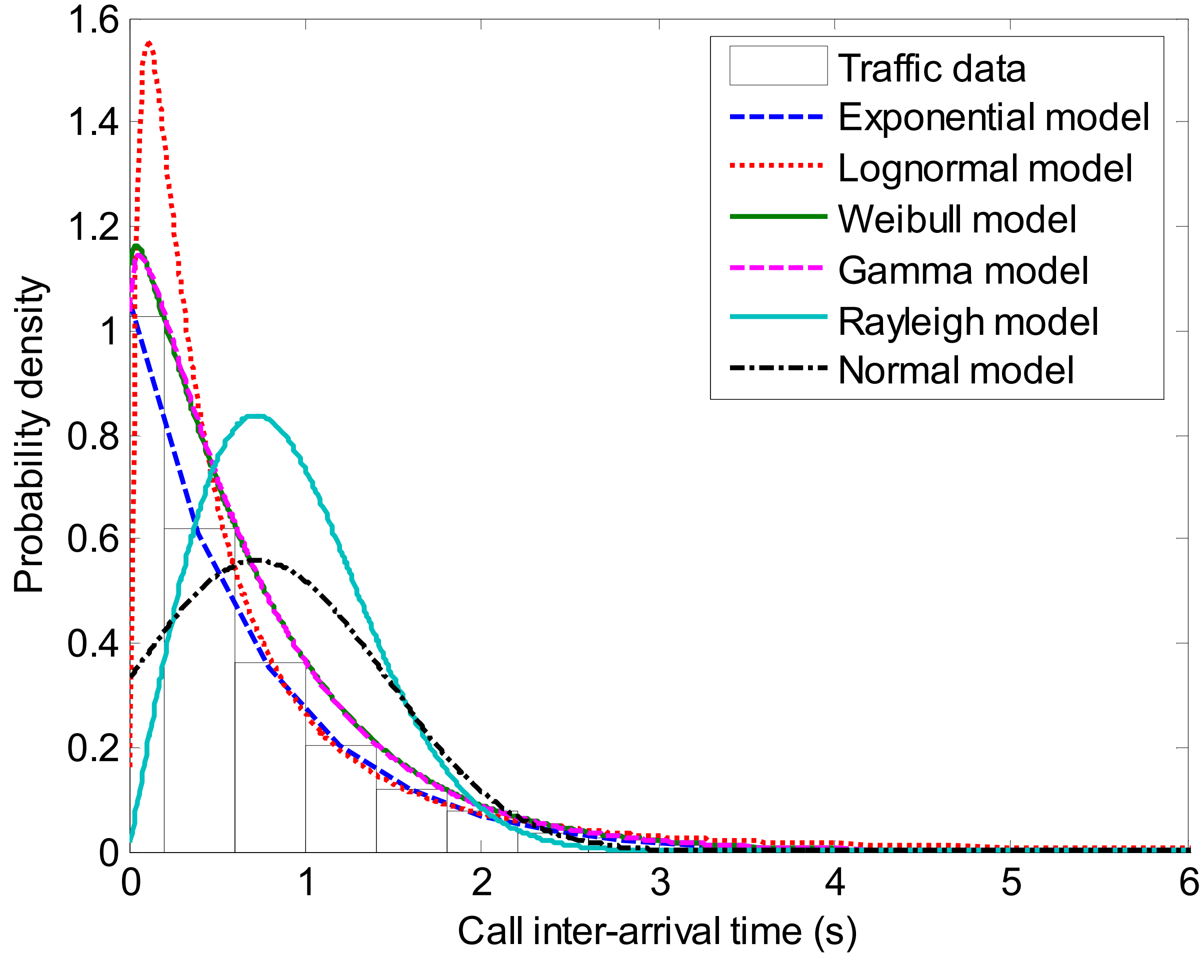
$$L(x_1, x_2, \ldots, x_N | \theta_1, \theta_2, \ldots, \theta_k) = L = \prod_{i=1}^{N} f(x_i; \theta_1, \theta_2, \ldots, \theta_k)$$

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

Maximum likelihood estimation



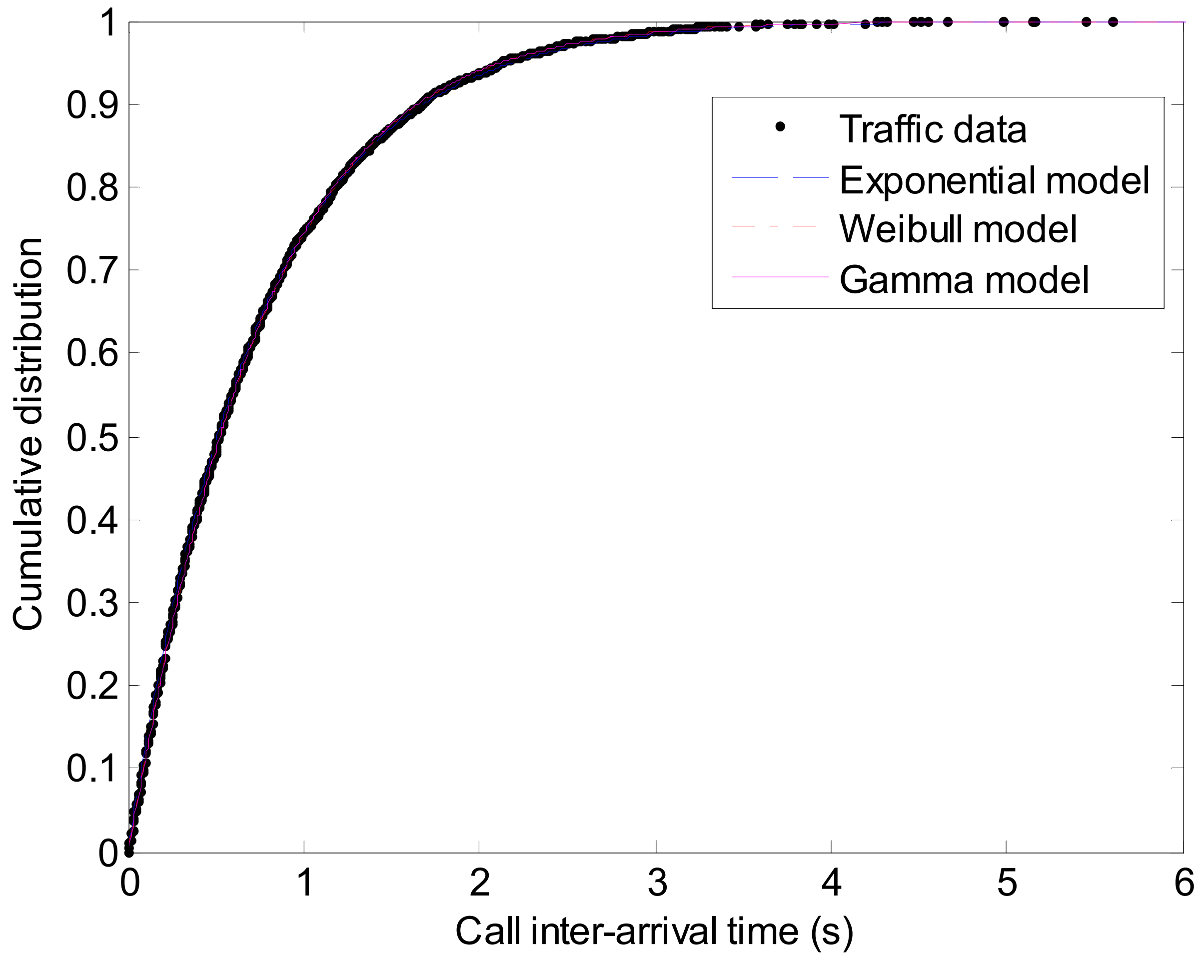
Call inter-arrival times: pdf candidates



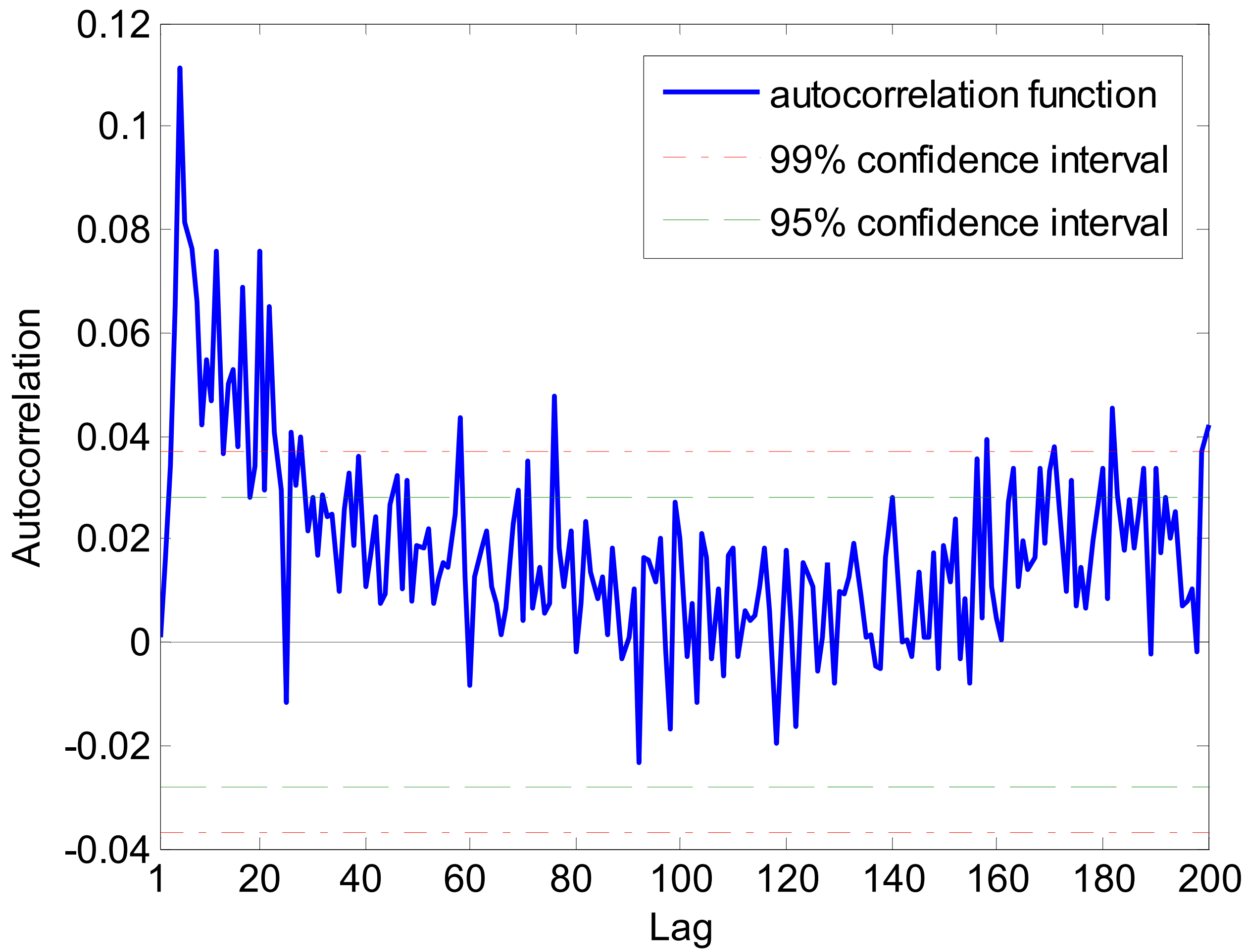
K-S test results: 2003

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

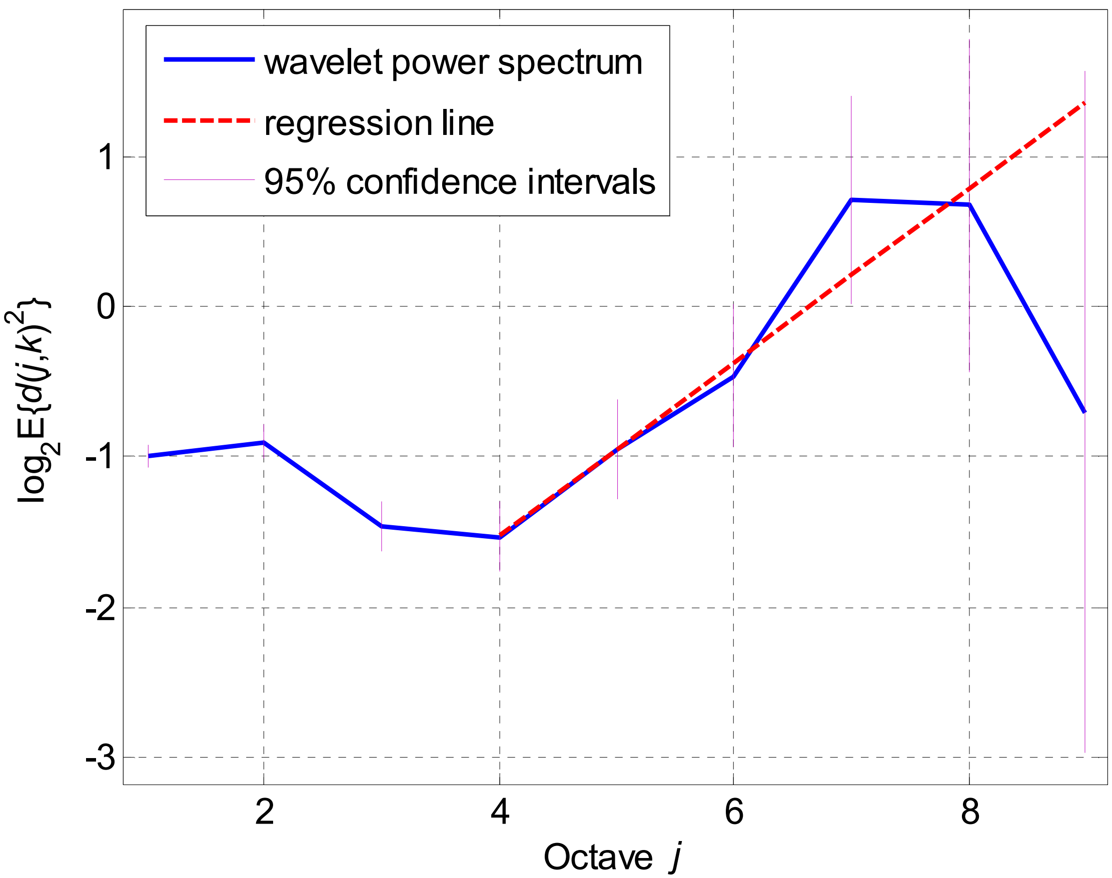
Call inter-arrival times, best-fitting distributions: cdf



Call inter-arrival time: autocorrelation



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



- LRD: α>0 (H>0.5)
- similar logscale diagrams for other traces

Call inter-arrival times: estimates of H

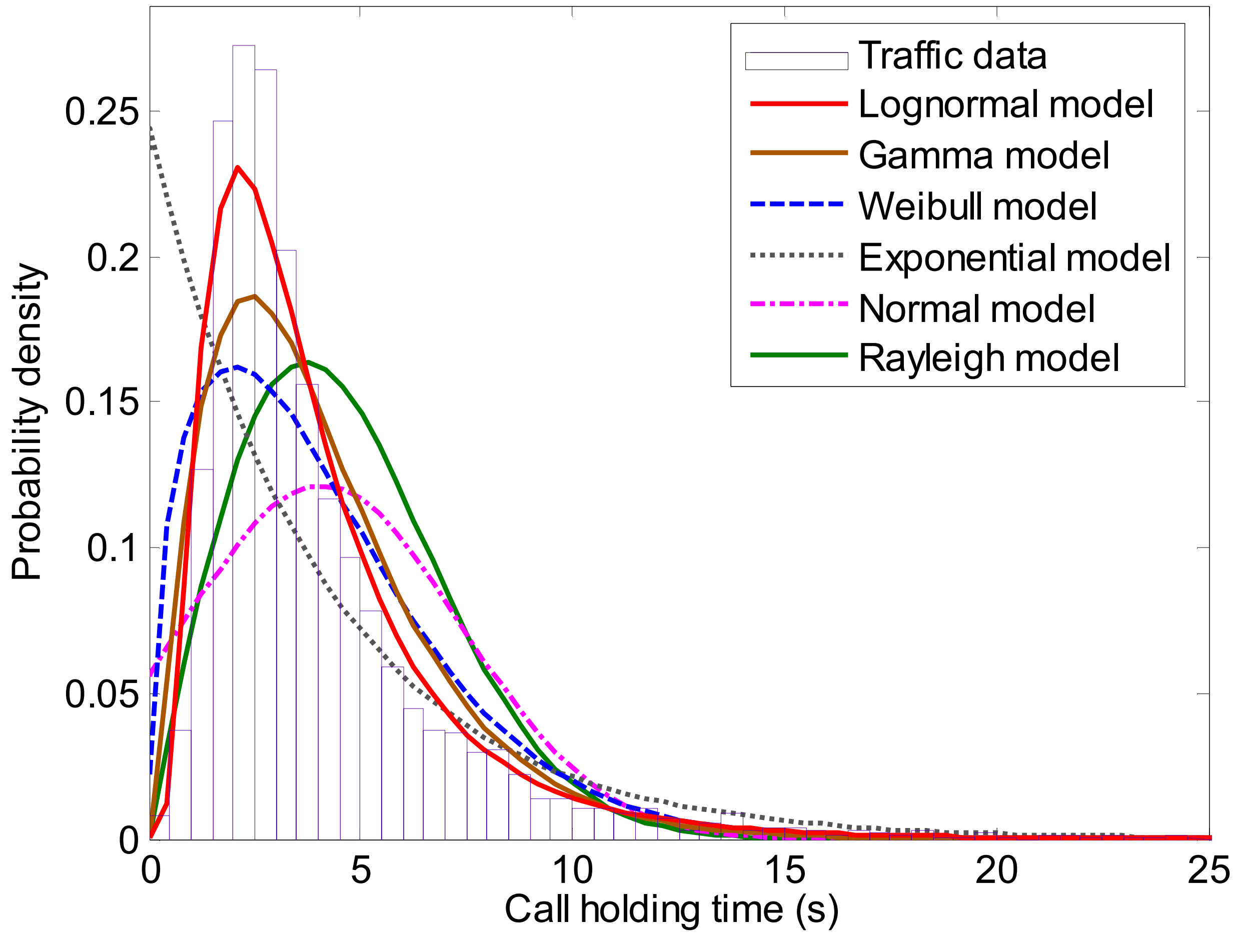
- Traces pass the test for time constancy of α: estimates of H are reliable

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

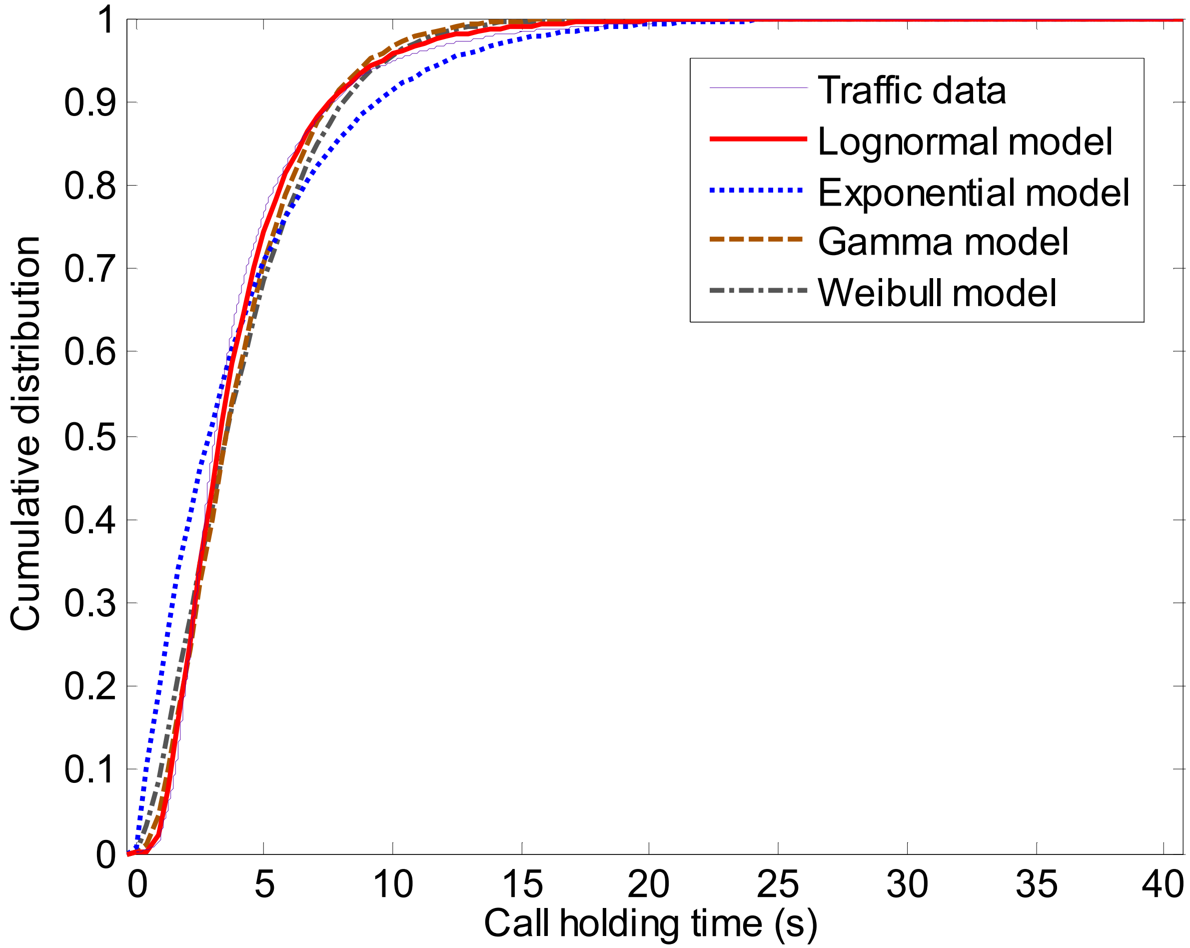
Roadmap

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

Call holding time: pdf candidates



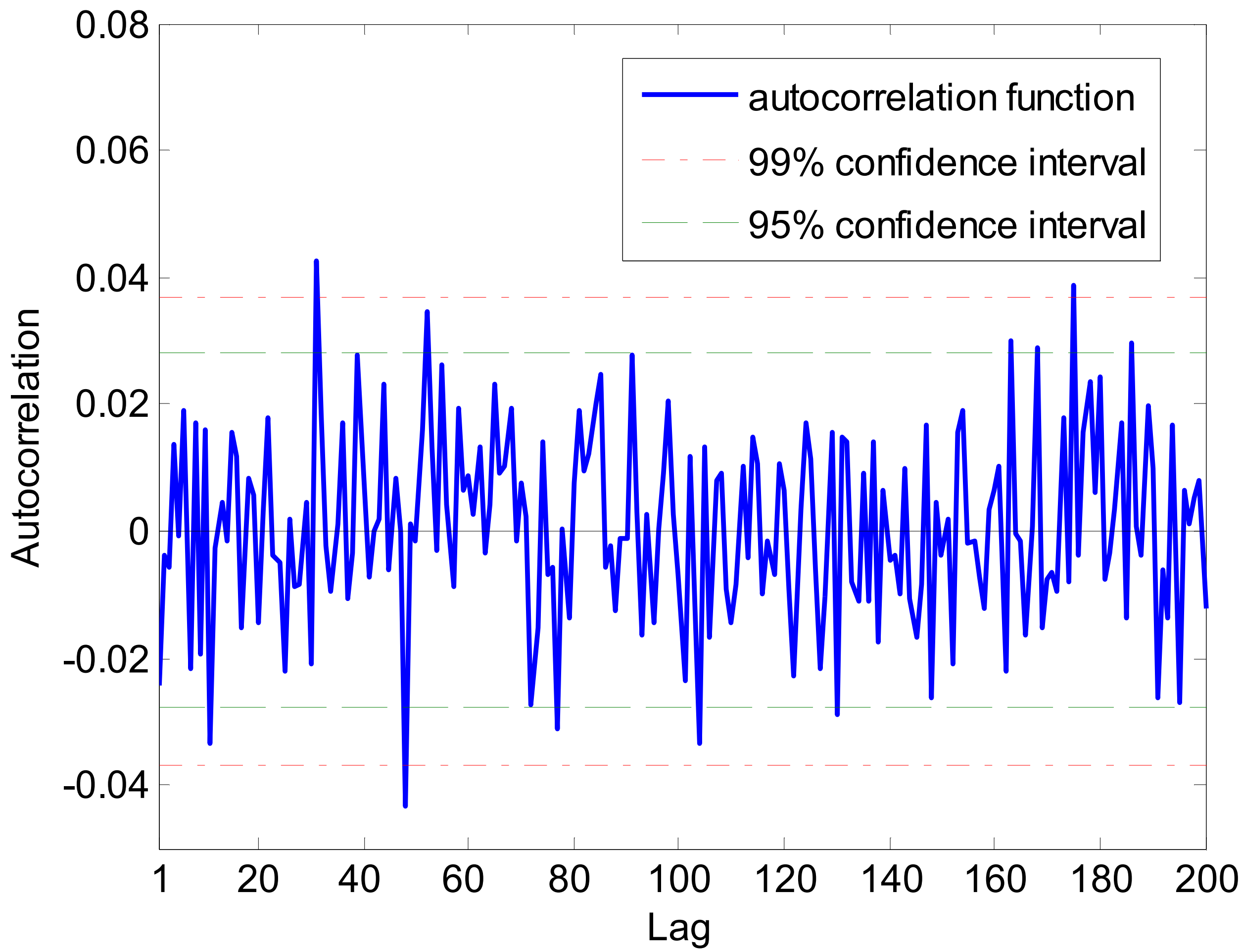
Best-fitting distributions: cdf



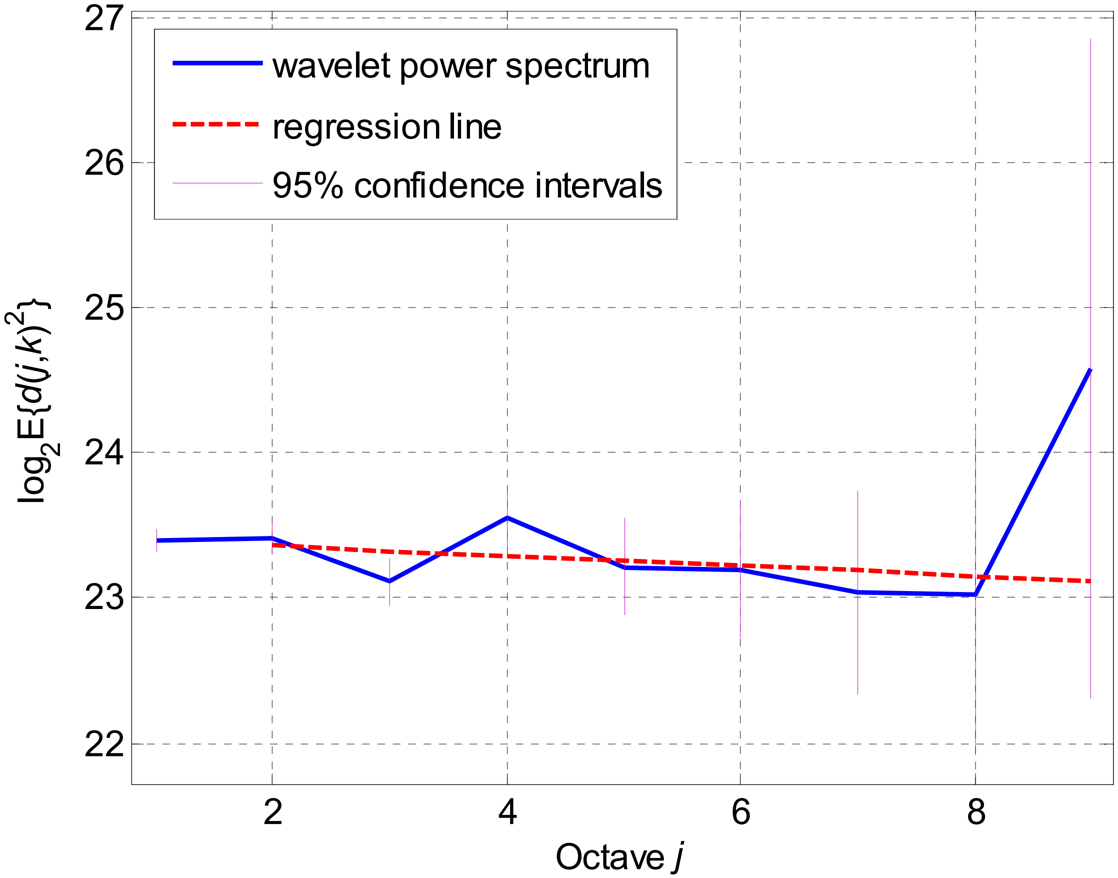
K-S test results: 2003

- 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

Roadmap

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

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)

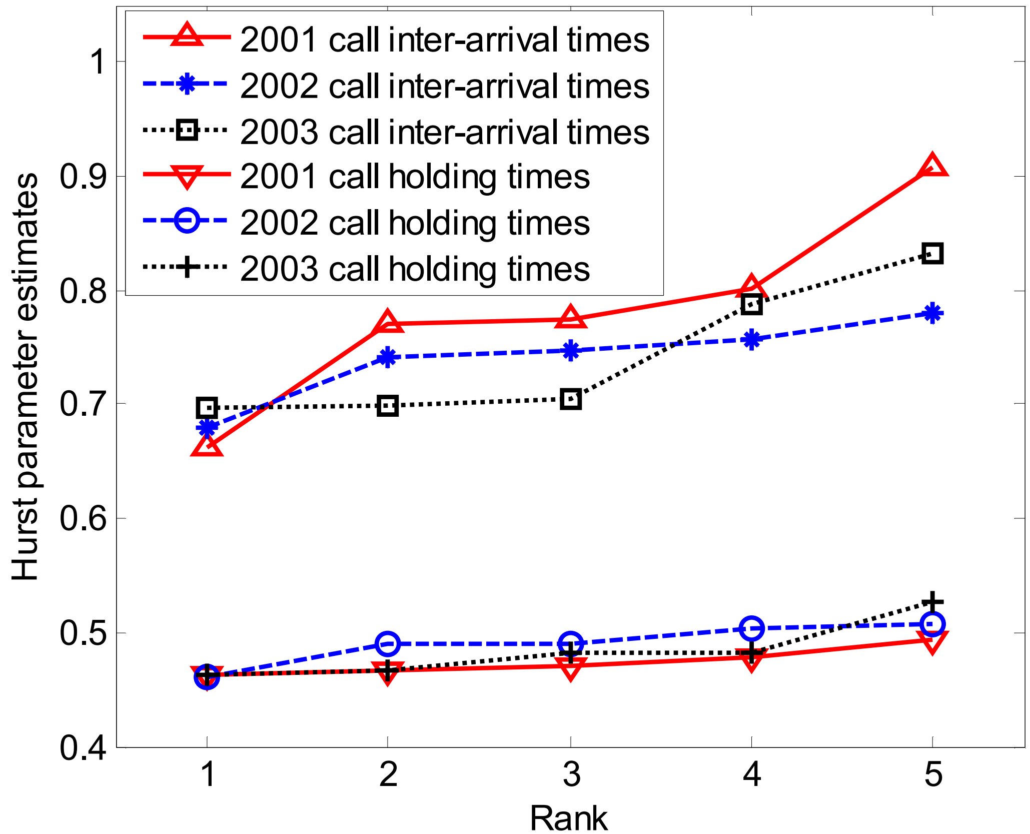
Distributions

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

Conclusions

- 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

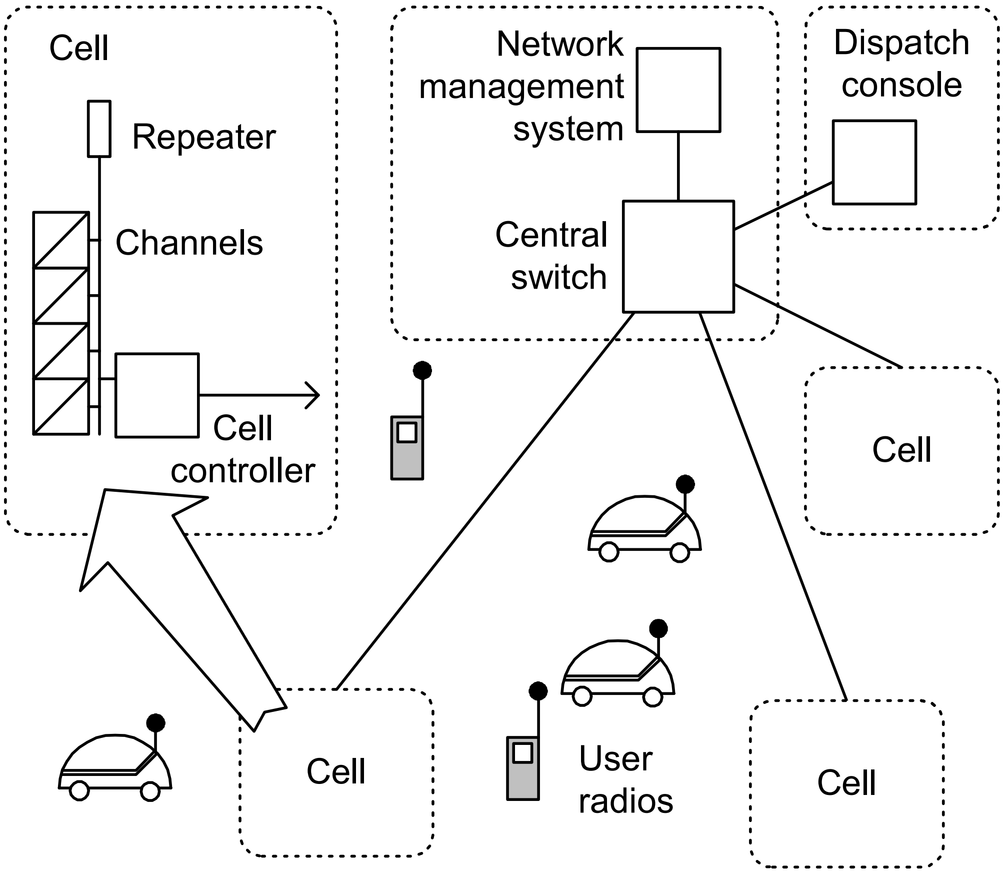
Conclusions

- 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

References

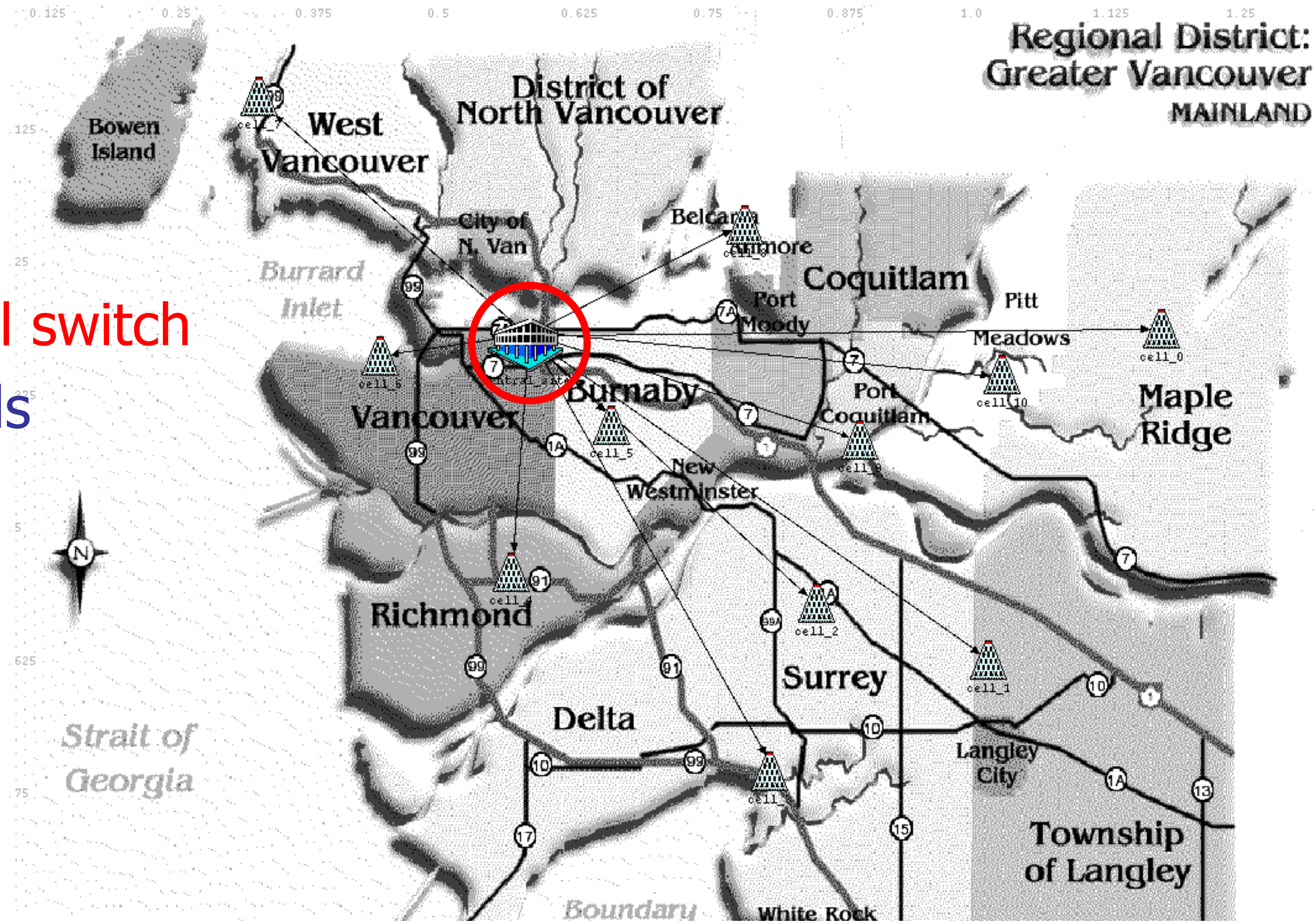
- J. Song and Lj. Trajković, "Modeling and performance analysis of public safety wireless networks," in *Proc. IEEE IPCCC*, Phoenix, AZ, Apr. 2005, pp. 567–572.
- D. Sharp, N. Cackov, N. Lasković, Q. Shao, and Lj. Trajković, "Analysis of public safety traffic on trunked land mobile radio systems," *IEEE J. Select. Areas Commun.*, vol. 22, no. 7, pp. 1197–1205, Sept. 2004.
- 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.
- E-Comm, Emergency Communications for SW British Columbia Incorporated. (2005, May). [Online]. Available: http://www.ecomm.bc.ca.
- 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.
- 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, Feb. 1994.
- 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.
- 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.

Network architecture

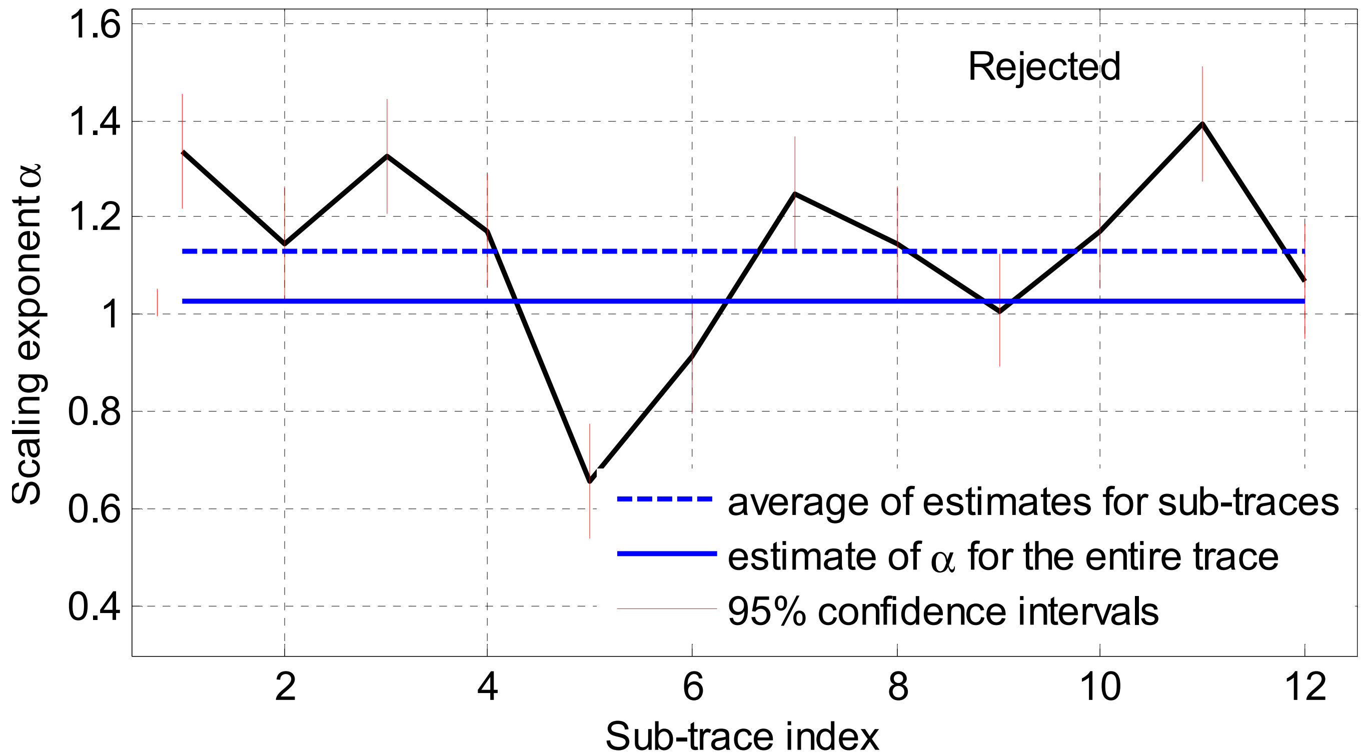


Network model

- central switch
- 11 cells



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)