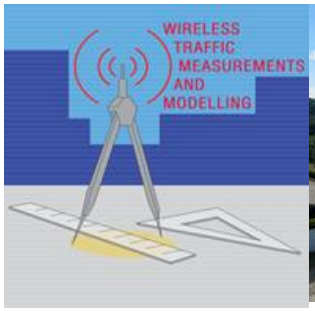
On Scalable Measurement-driven Modeling of Traffic Demand in Large WLANs



Prof. Maria Papadopouli [1,2]

Merkouris Karaliopoulos[2] Haipeng Shen[2] Elias Raftopoulos[1]

[1] Foundation for Research & Technology-Hellas (FORTH) & University of Crete
[2] University of North Carolina at Chapel Hill

[1]IBM Faculty Award, EU Marie Curie IRG, GSRT grants

Wireless landscape

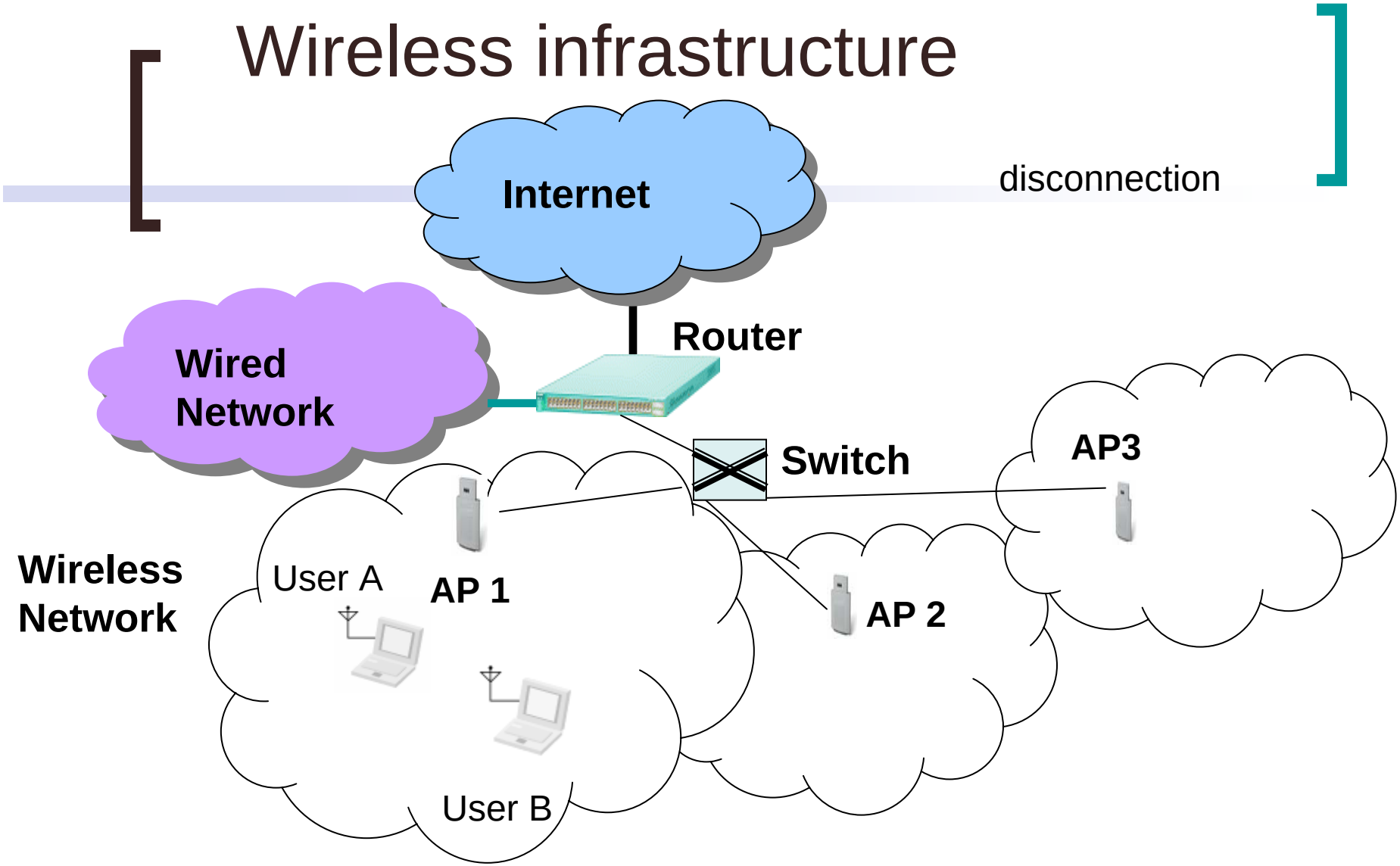
- Growing demand for wireless access
- Mechanisms for better than best-effort service provision

  Admission control, channel switching, load balancing, roaming
- Performance analysis of these mechanisms
- Majority of studies make high-level observations about traffic dynamics in tempo-spatial domain
- Models of network & user activity in various spatio-temporal scales are required

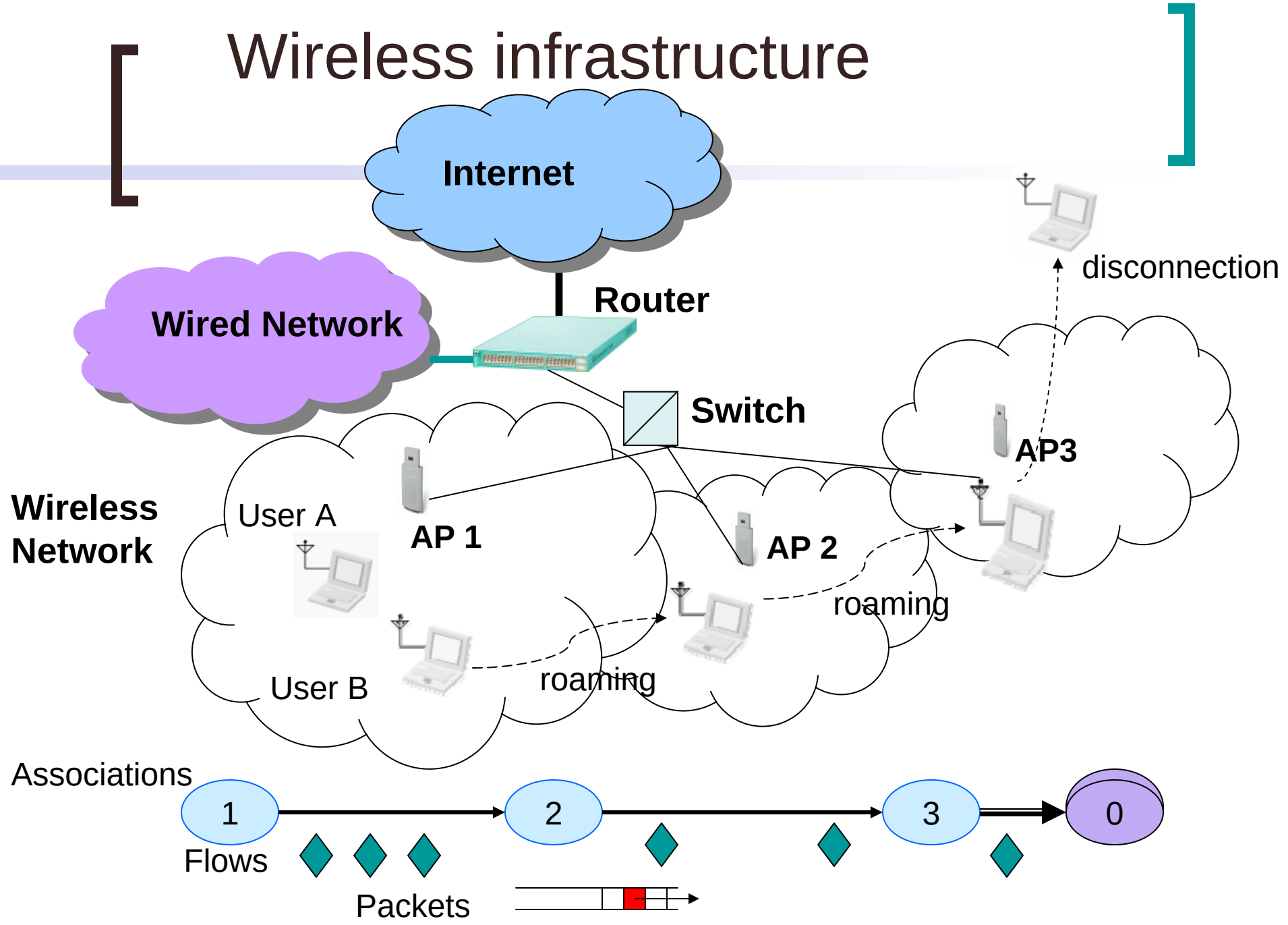
Roadmap

- Wireless infrastructure and access
- Modeling objectives & structures
- Related research
- Main research issues
- Validation of models
- Conclusions & future work

Wireless infrastructure



Wireless infrastructure

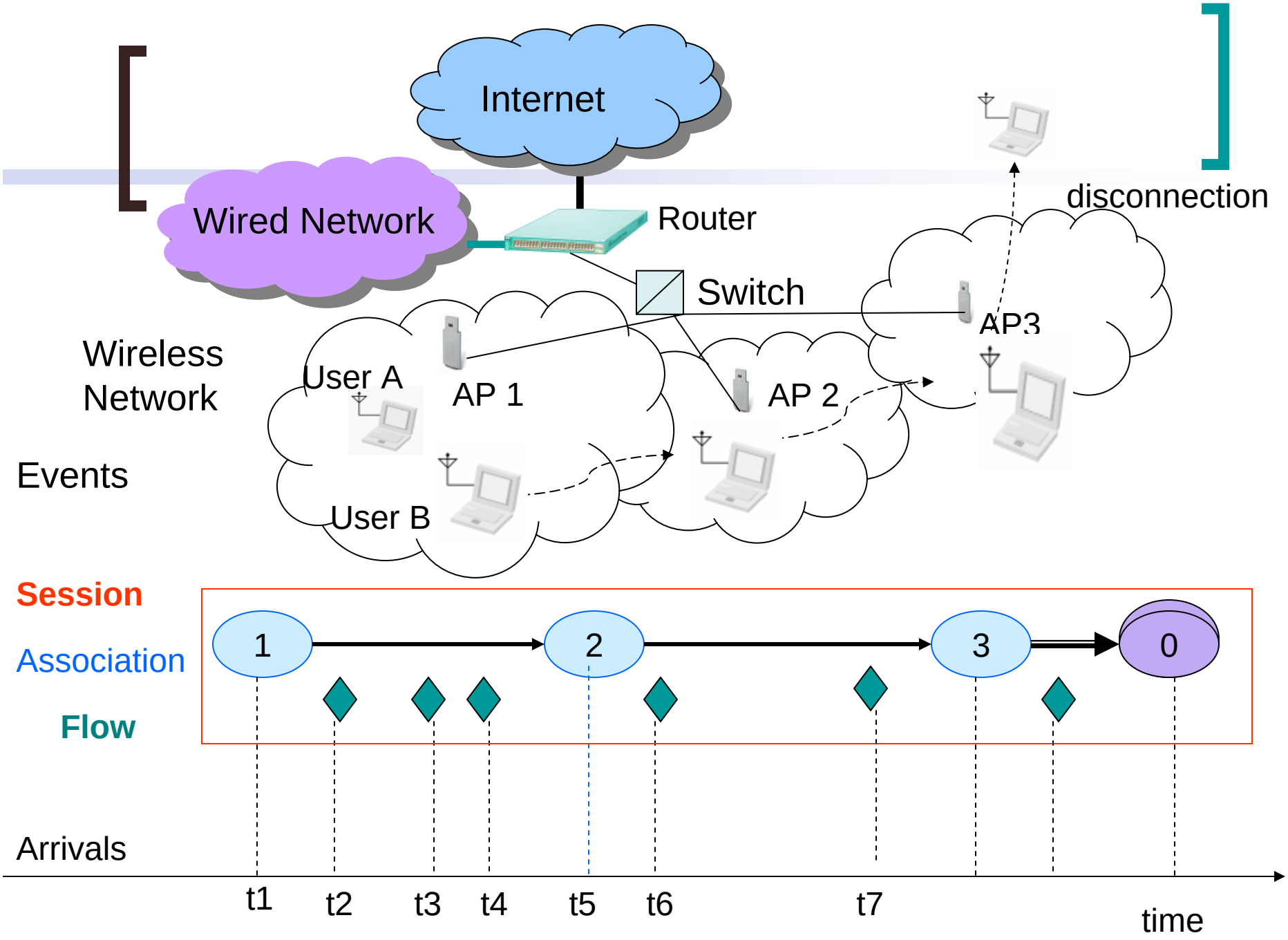


Roadmap

- Wireless infrastructure and access
- Modeling objectives & structures
- Related research
- Main research issues
- Validation of models
- Conclusions & future work

Modelling objectives

- Important dimensions on wireless network modelling
  - user demand (access & traffic)
  - topology (network, infrastructure, radio propagation)
- Structures that are well-behaved, robust, scalable & reusable
- Publicly available analysis tools, traces, & models


Internet
Wired Network
Router
disconnection
Switch
Wireless
Network
User A
AP 1
AP 2
AP3
User B
Events
Session
Association
Flow
1
2
3
0
Arrivals
t1
t2
t3
t4
t5
t6
t7
time

Wireless infrastructure & acquisition

- 26,000 students, 3,000 faculty, 9,000 staff in over 729-acre campus
- 488 APs (April 2005), 741 APs (April 2006)
- SNMP data collected every 5 minutes
- Packet-header traces:
  - 175GB (April 2005), 365GB (April 2006)
  - captured on the link between UNC & rest of Internet via a high-precision monitoring card

Main modeling structures

- Session
  - arrival process
  - starting AP

    Captures interaction between clients & network

- Flow within session
  - arrival process
  - number of flows
  - size

Above packet level for traffic analysis & closed-loop traffic generation

Roadmap

- Wireless infrastructure and access
- Modeling objectives & structures
- Related research
- Main research issues
- Validation of models
- Conclusions & future work

Related research in wired networks

- Flow-level
  - several protocols (mainly TCP)
- Session-level
  - FTP, web traffic
  - session borders heuristically defined by intervals of inactivity

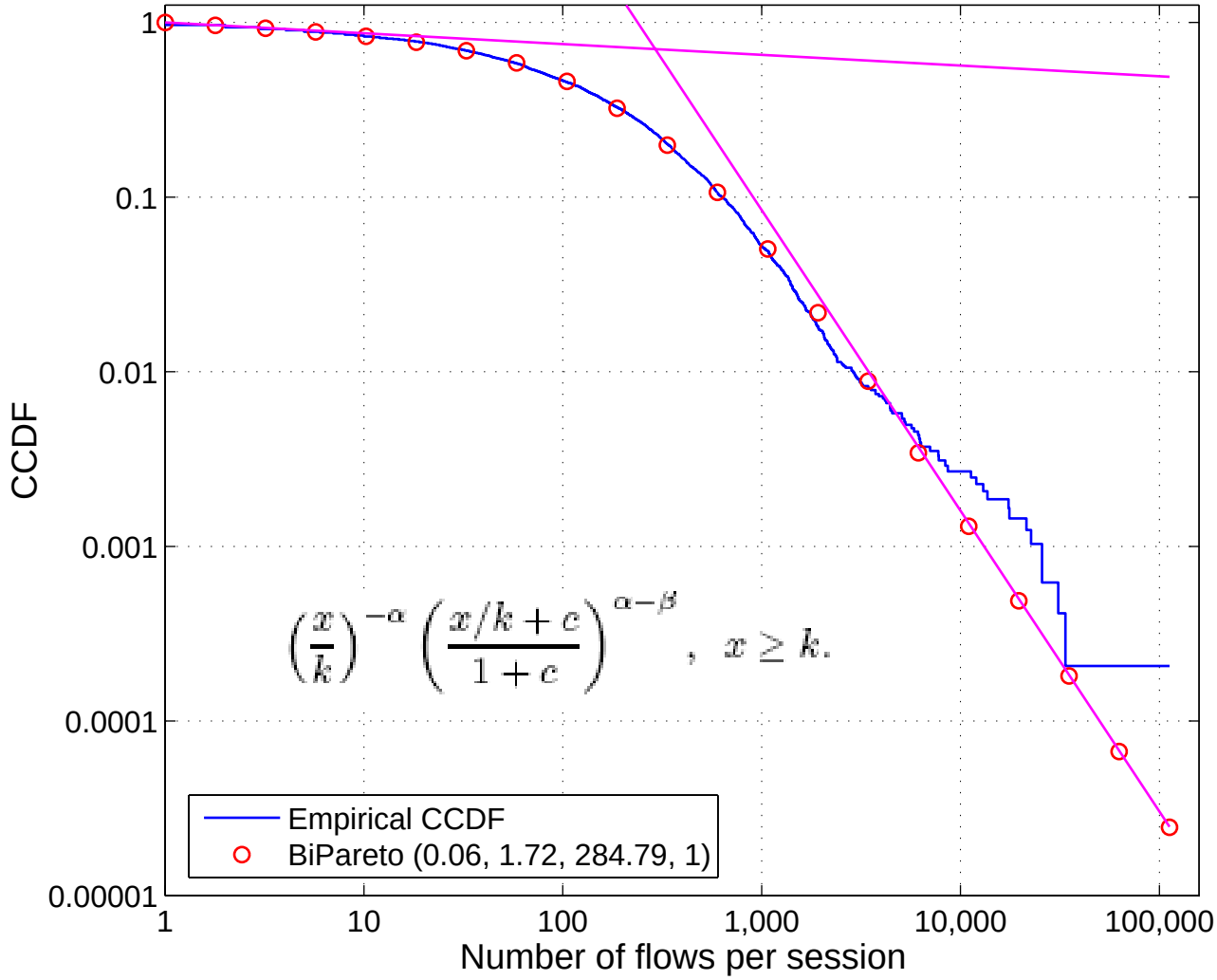
Related research in wireless networks

- Flow-level modeling by Meng [mobicom '04]
  - No session concept,
  - Weibull for flow interarrivals
  - AP-level over hourly intervals

- Hierarchical modeling by Papadopouli [wicon '06]
  - Parameters: Session & in-session flow:
  - Time-varying Poisson process for session arrivals
  - BiPareto for in-session flow numbers & flow sizes
  - Lognormal for in-session flow interarrivals

☞ Reduces the deviation from real traces

Number of Flows Per Session



Roadmap

- Wireless infrastructure and access
- Modeling objectives & structures
- Related research
- Main research issues
- Validation of models
- Conclusions & future work

Tradeoffs in these modeling approaches

Scale \ Objective	Accuracy	Scalability	Amenability to analysis
Hourly period @ AP			
Network-wide			

Main research issues

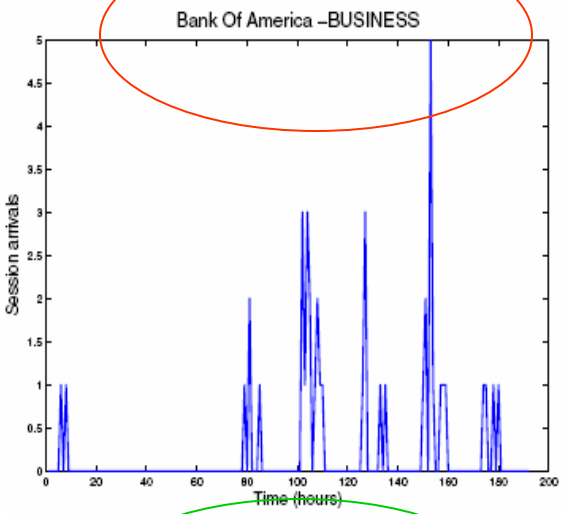
- Hierarchical modeling traffic workload

  AP-level vs. network-wide

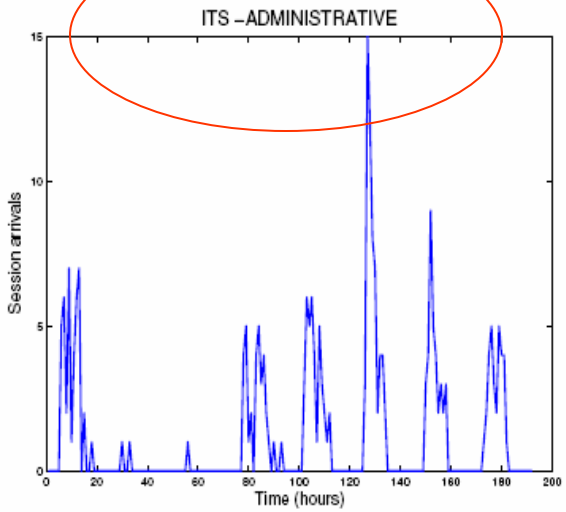
  Other spatio-temporal levels ?

- Model validation @ different spatial scales using data from different periods
- Scalability, reusability, accuracy tradeoffs

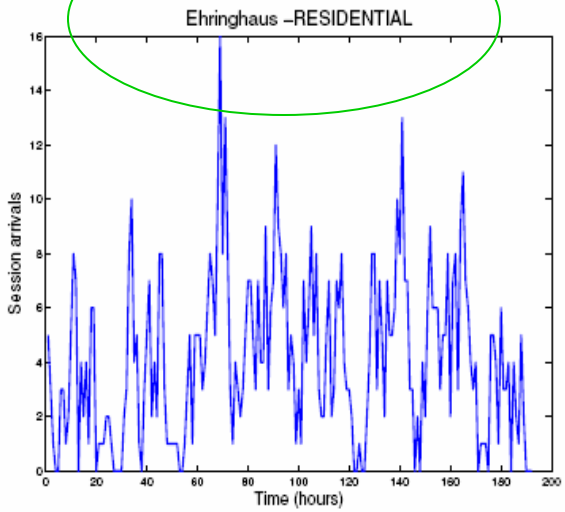
Hourly session arrival rates



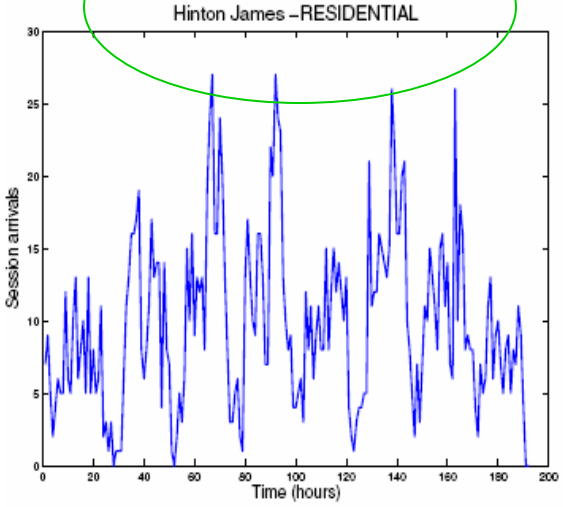
(a)

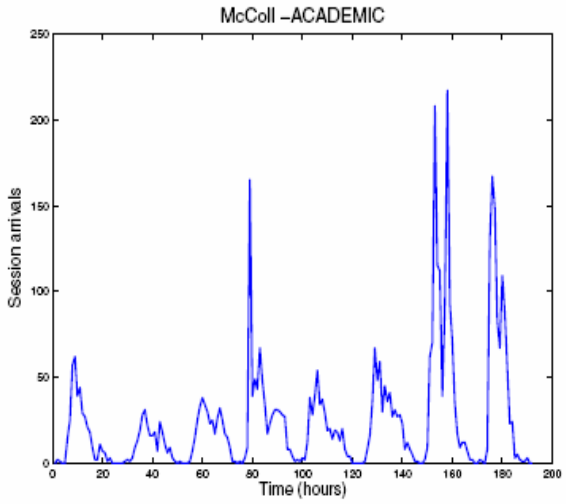


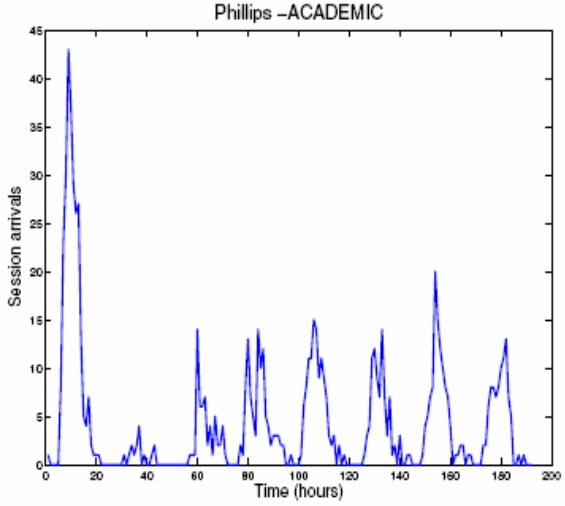
(b)



(c)

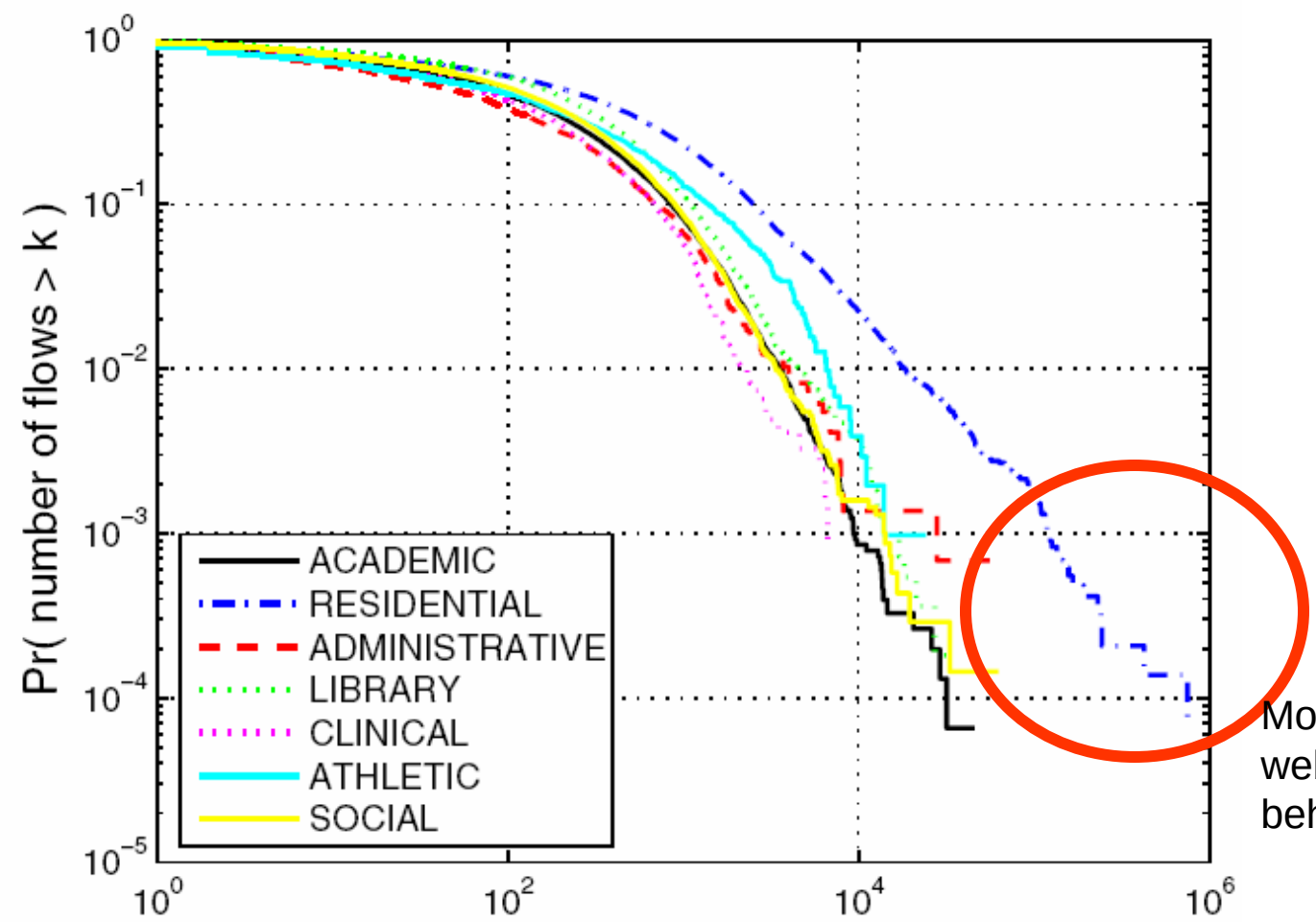






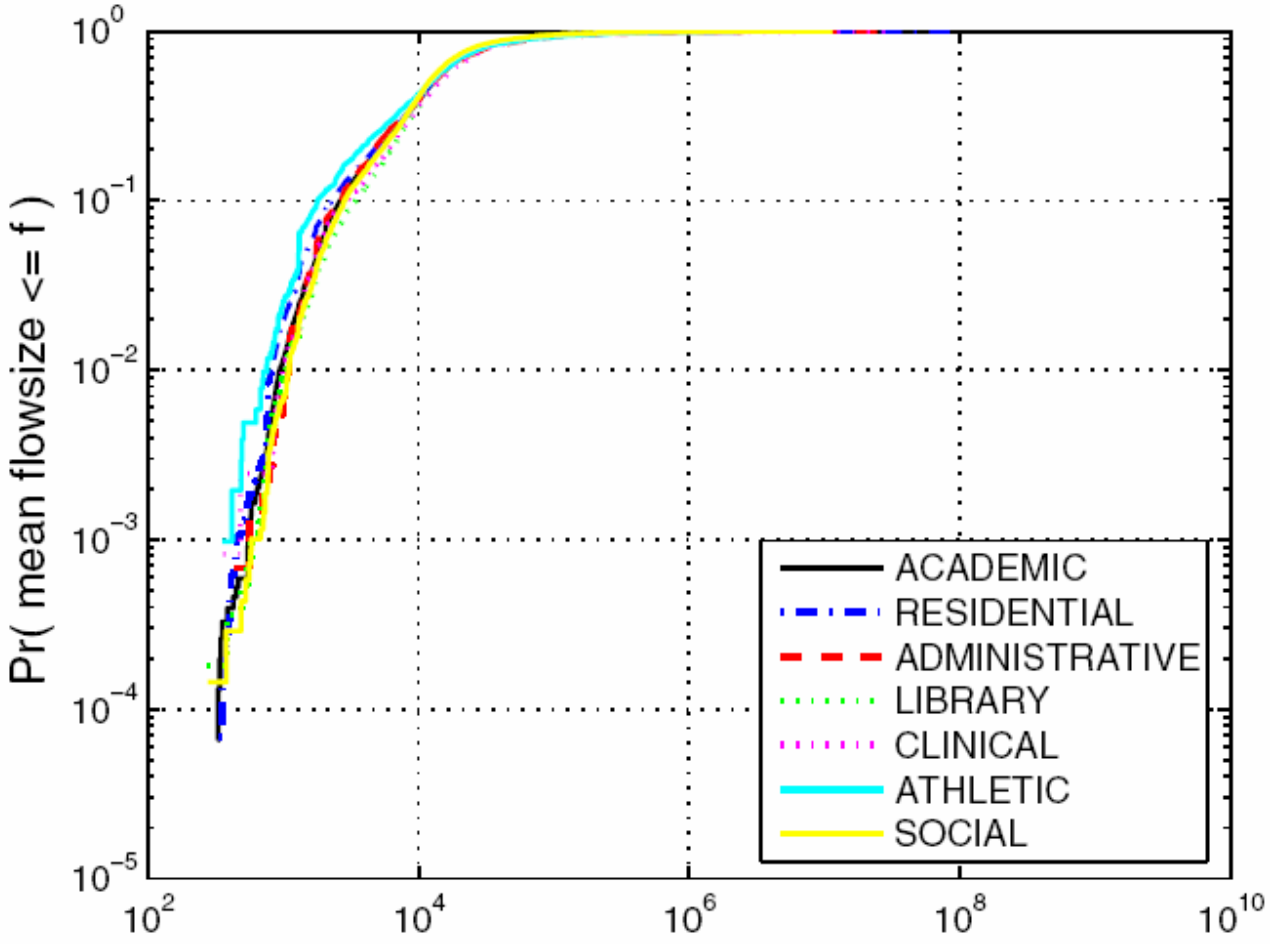
Number of flows in-session

Broad variation of the in-session number of flows per building-type distribution



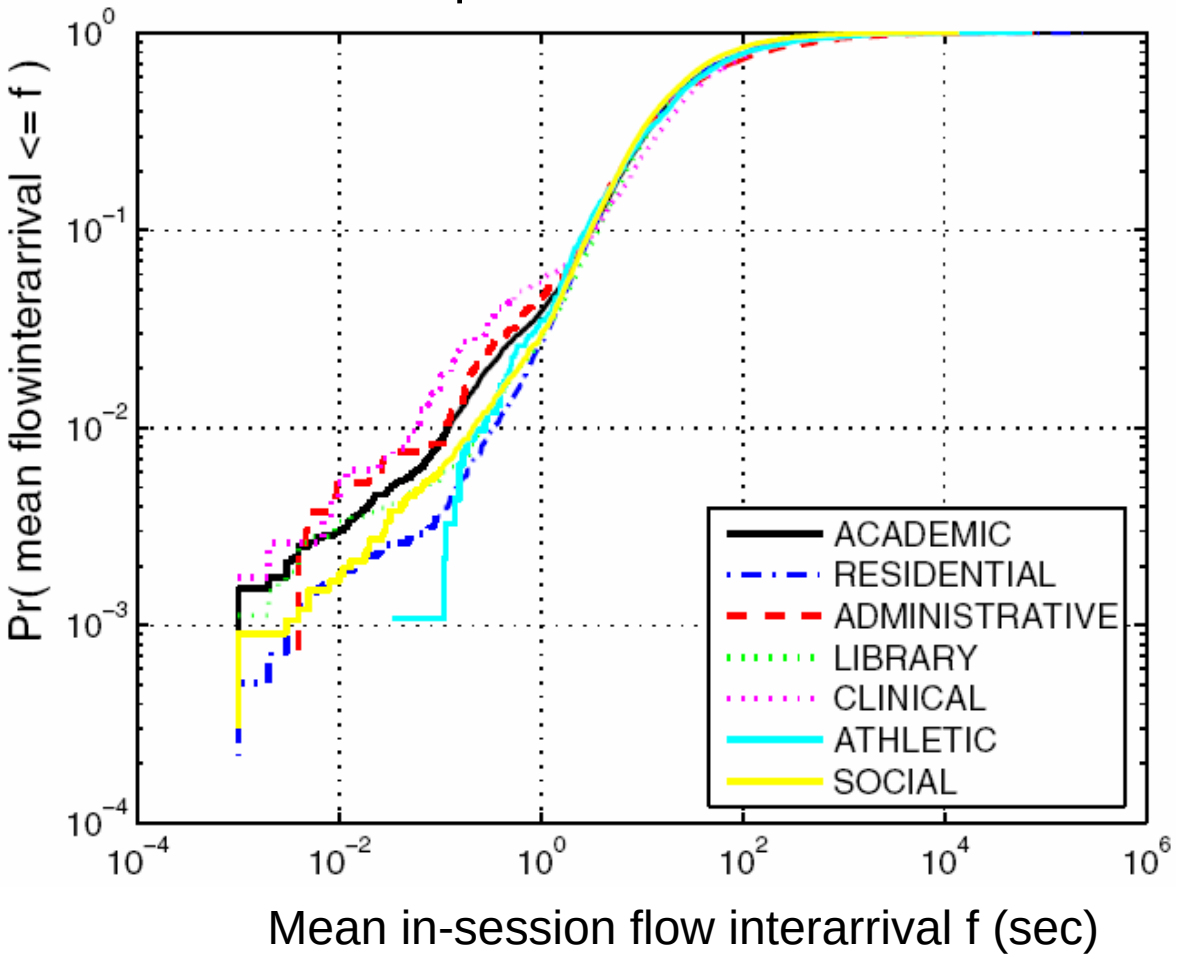
More active web browsing behavior

Session-level flow size



Session-level flow related variation

In-session flow interarrival can be modeled with same distribution for all building types but with different parameters



Roadmap

- Wireless infrastructure and access
- Modeling objectives & structures
- Related research
- Main research issues
- Validation of models
- Conclusions & future work

Model validation

- ☞ Compare synthetic vs. original trace

- Metrics: variables not explicitly addressed by our models
  - aggregate flow arrival count process
  - aggregate flow interarrival (1st & 2nd order statistics)

- Increasing order of spatial aggregation

  AP-level, building-level (bldg), building-type-level (bldg-type), network-wide

- Different tracing periods (April 2005 & 2006)

Synthetic traces

- Synthesize sessions & flows based on aforementioned distributions

- Parameters estimation
  - Session arrival

    using real hourly bldg-specific data
  - Flow interarrival & number of flows

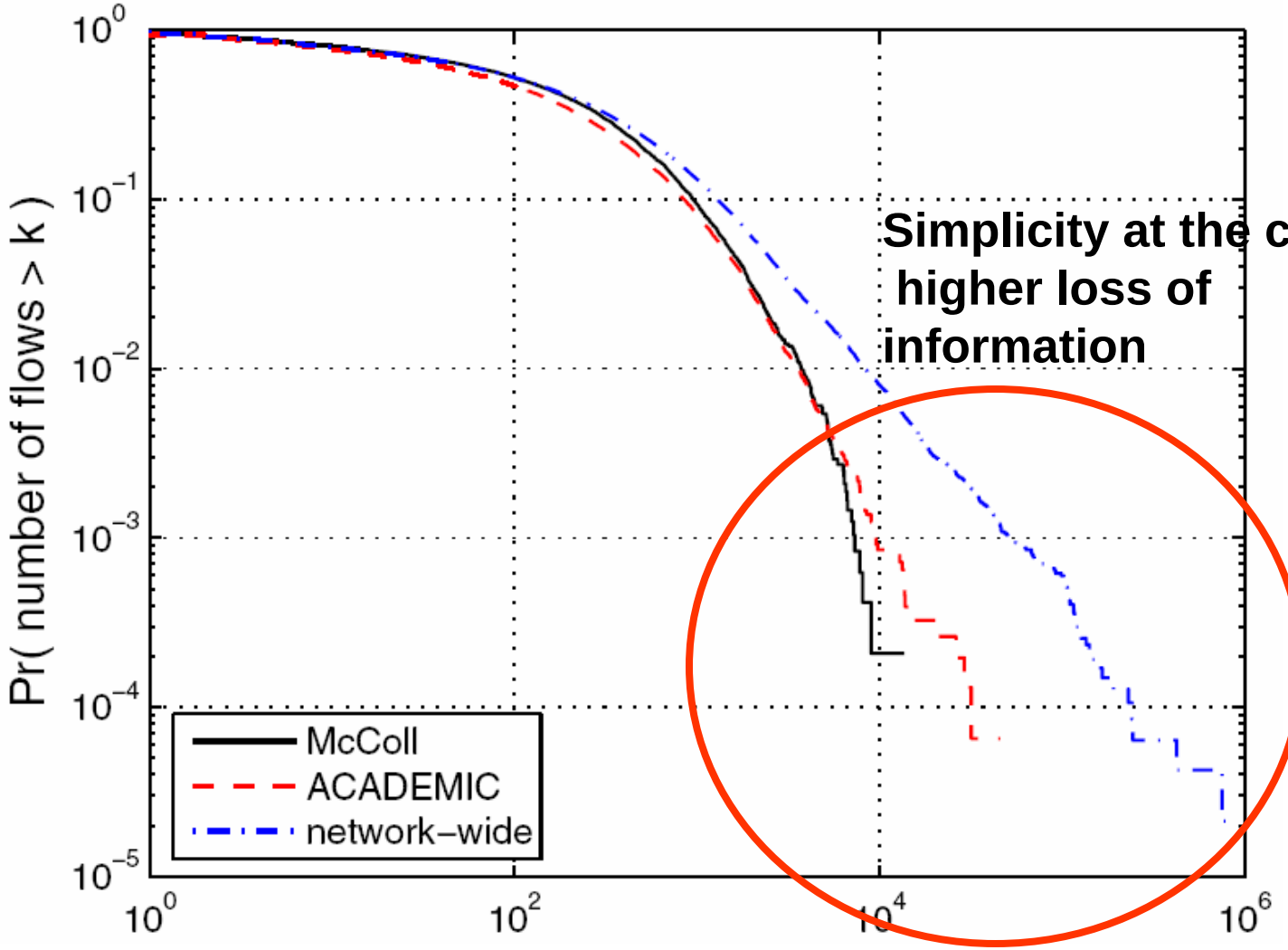
    depending on scale using the respective real data

    bldg (day), bldg (trace), bldg-type, network-wide

    Notation: Session–Flow(duration of trace)

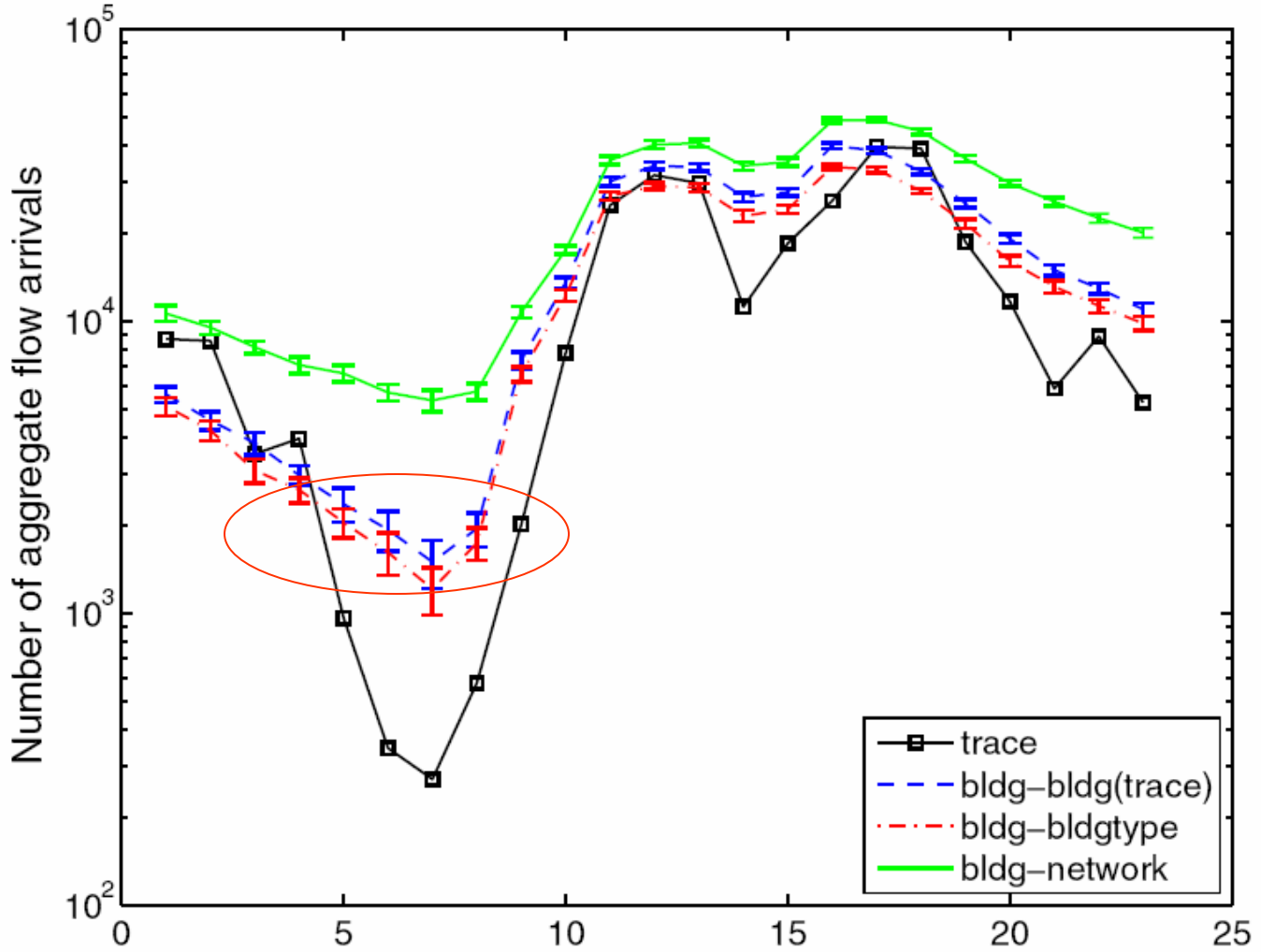
    Example: bldg–bldg(day)

Number of flows per session

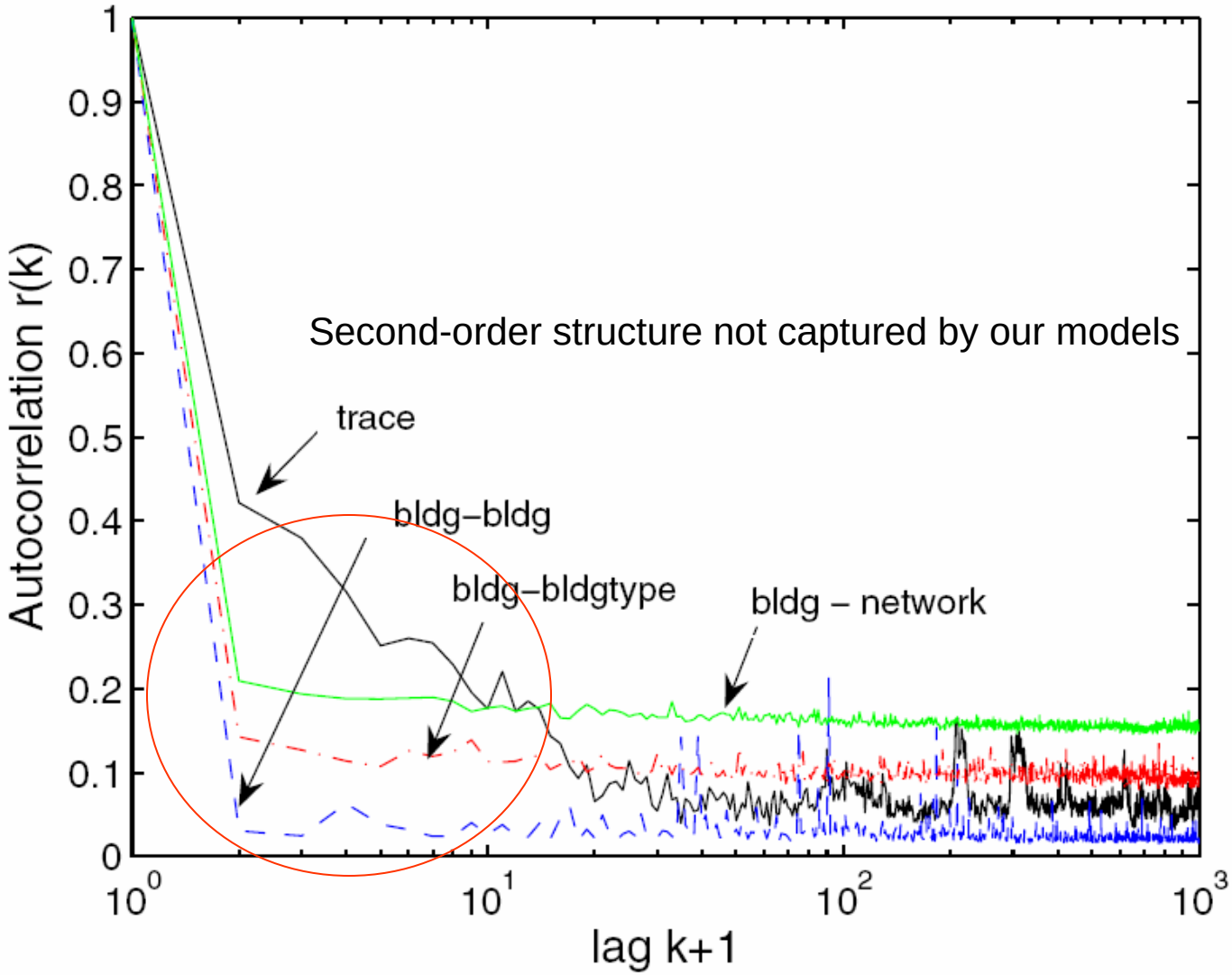


Simplicity at the cost of higher loss of information

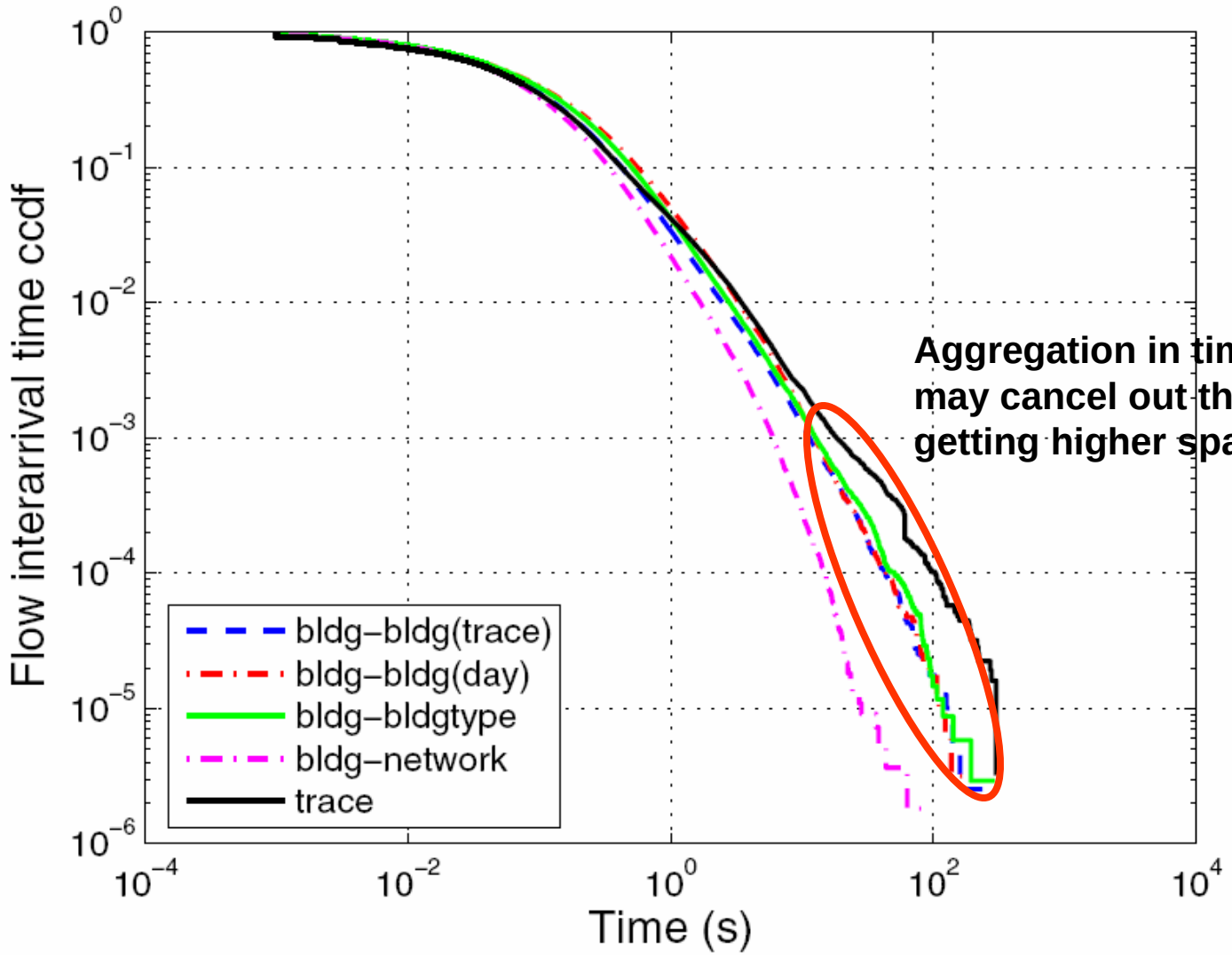
Number of aggregate flow arrivals



Autocorrelation of flow interarrivals



Flow interarrivals time



Aggregation in time-dimension may cancel out the benefit of getting higher spatial resolution

Conclusions

Multi-level parametric modelling of wireless demand

- Network-wide models:
  - Time-varying Poisson process for session arrivals
  - biPareto for in-session flow numbers & flow sizes
  - Lognormal for in-session flow interarrivals

✓ Validation of models over two different periods

✓ Same distributions apply for modeling at finer spatial scales building-level, groups of buildings with similar usage

✓ Evaluation of scalability-accuracy tradeoff

UNC/FORTH web archive

Online repository of models, tools, and traces

- Packet header, SNMP, SYSLOG, signal quality

  http://netserver.ics.forth.gr/datatraces/

Free login/ password to access it

Joint effort of Mobile Computing Groups @ FORTH & UNC

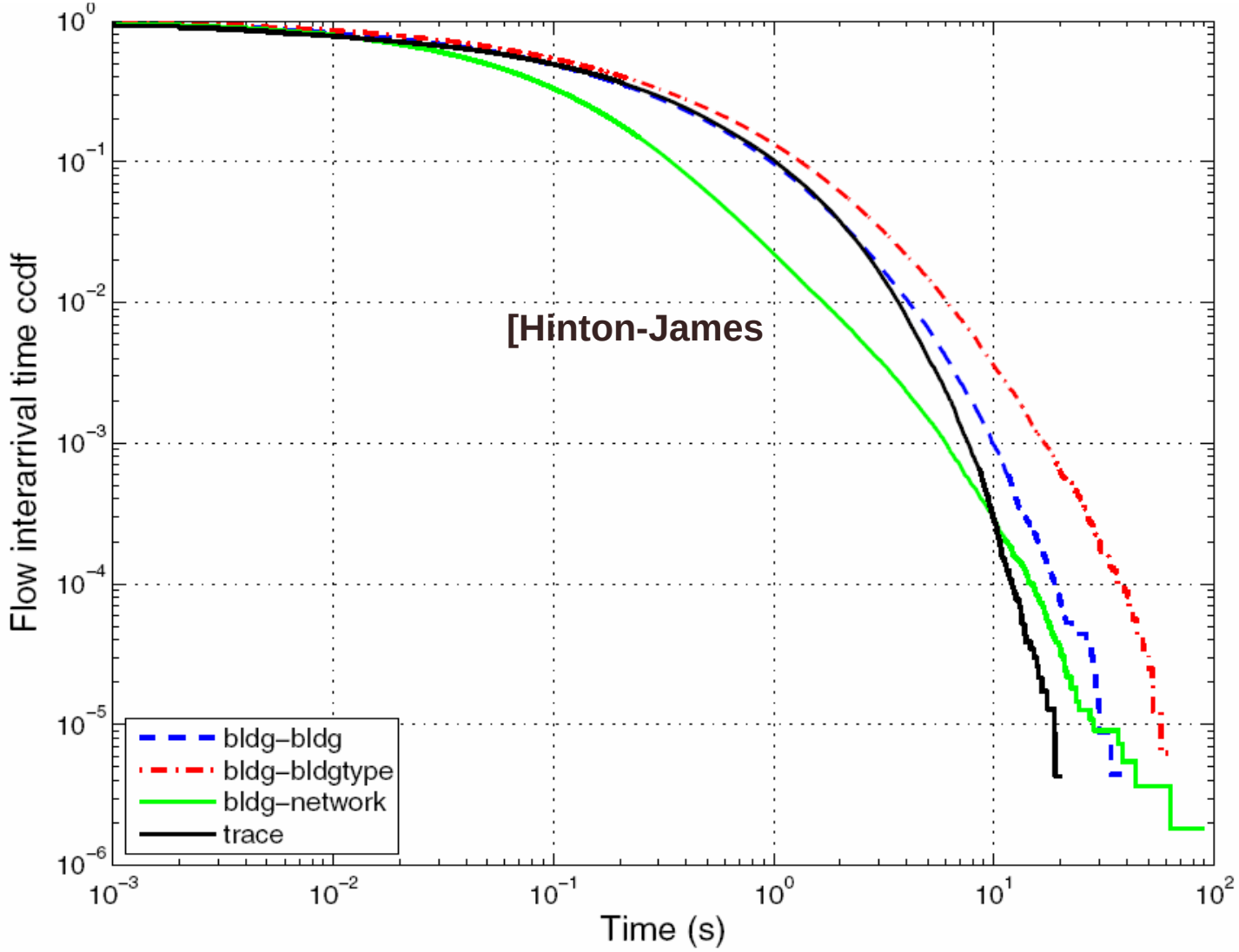
maria@csd.uoc.gr

Our parameters and models

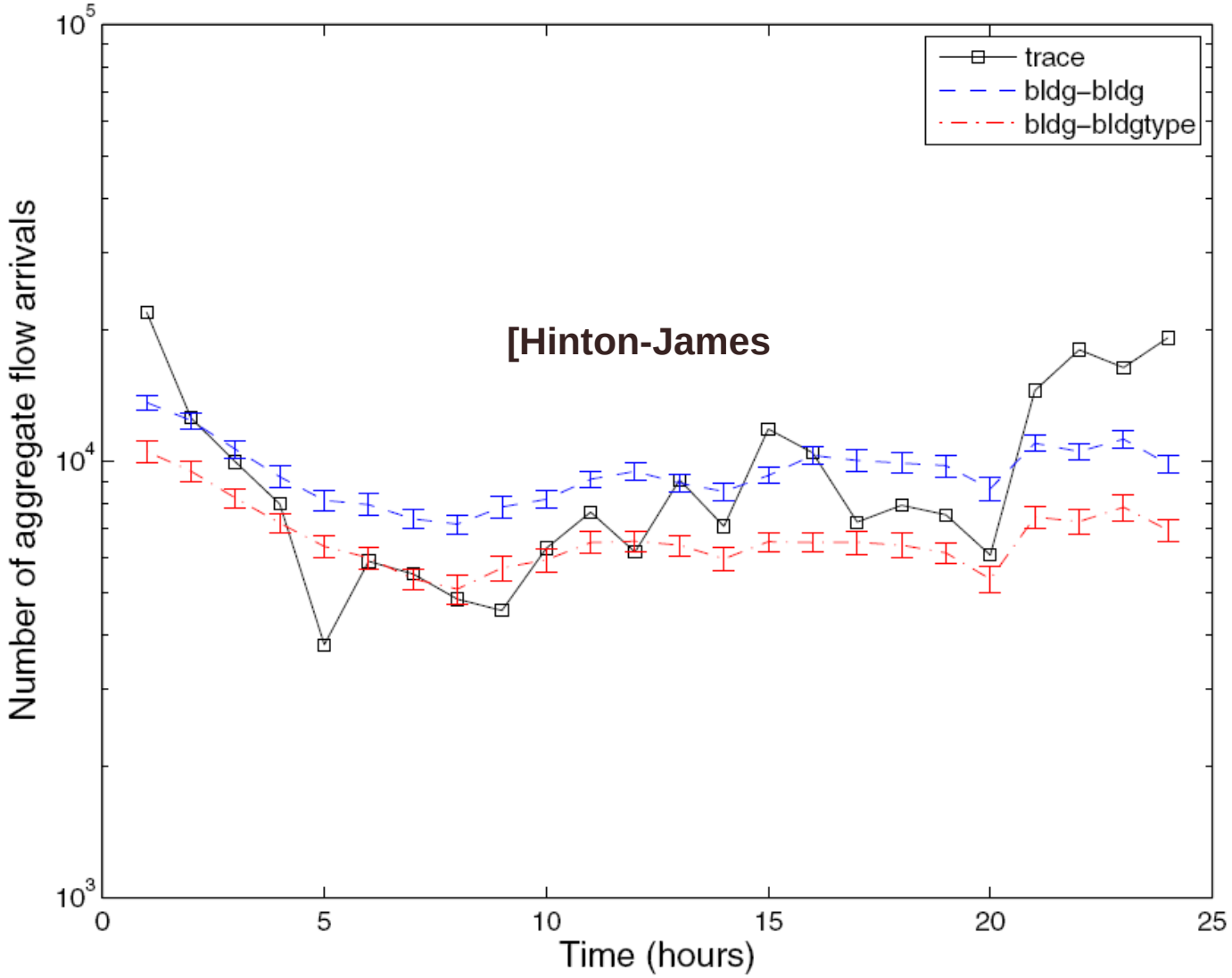
Parameter	Model	Probability Density Function	Related Papers
Association, session duration	BiPareto	$p(x) = k^b(1+c)^{b-a}x^{-a-1}(x+kc)^{a-b-1}(bx+akc), x>a$	EW' 06
Session arrival	Time-varying Poisson with rate λ(t)	N: # of sessions between t_1 and t_2 $\lambda = \int_{t_1}^{t_2}\lambda(t)dt,\ Pr(N=n) = \frac{e^{-\lambda}\lambda^n}{n!},\ n=0,1,\ldots$	WICON '06
Client arrival	Time-varying Poisson with rate λ(t)	Same as above	LANMAN '05
AP of first association/session	Lognormal	$p(x) = \frac{1}{\sqrt{2\pi}x\sigma}\exp\left[-\frac{(\ln x-\mu)2}{2\sigma 2}\right]$	WICON '06
Flow interarrival/session	Lognormal	Same as above	WICON '06
Flow number/session	BiPareto	$p(x) = k^b(1+c)^{b-a}x^{-a-1}(x+kc)^{a-b-1}(bx+akc), x>a$	WICON '06
Flow size	BiPareto	Same as above	WICON '06
Client roaming between APs	Markov-chain		INFOCOM'04
Spatio-temporal phenomena in wireless Web access			INFOCOM'04

Appendix

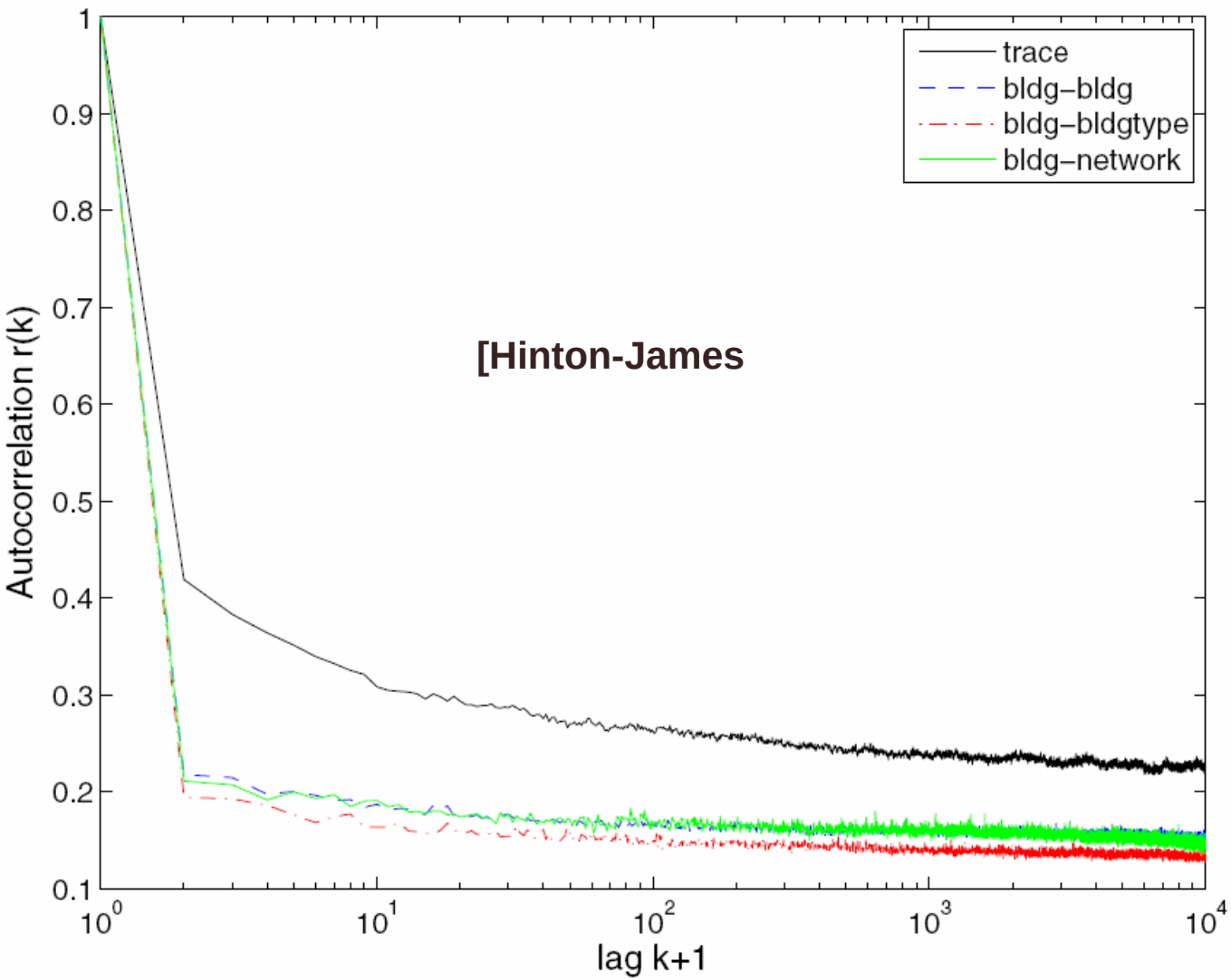
Flow interarrival time

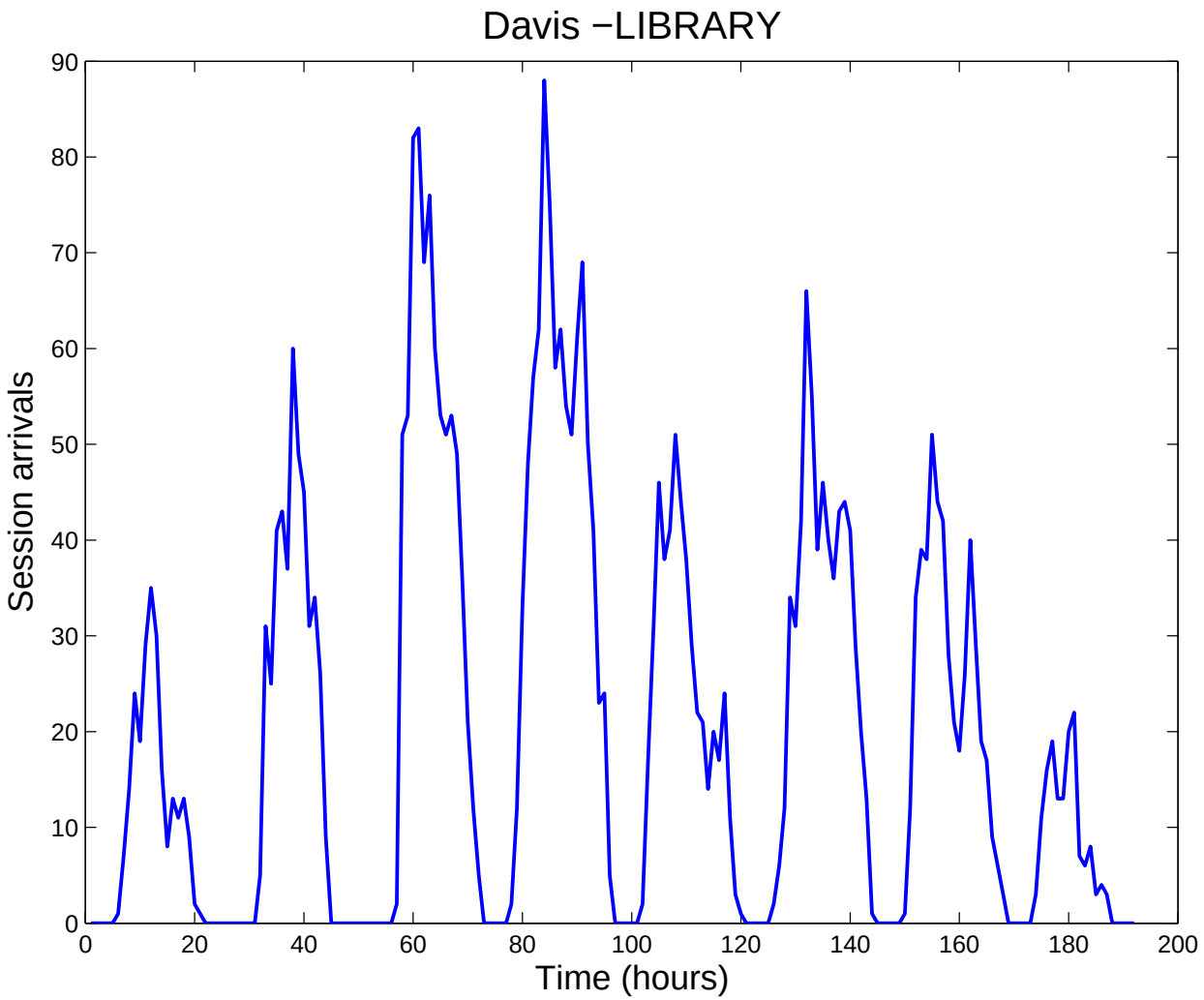


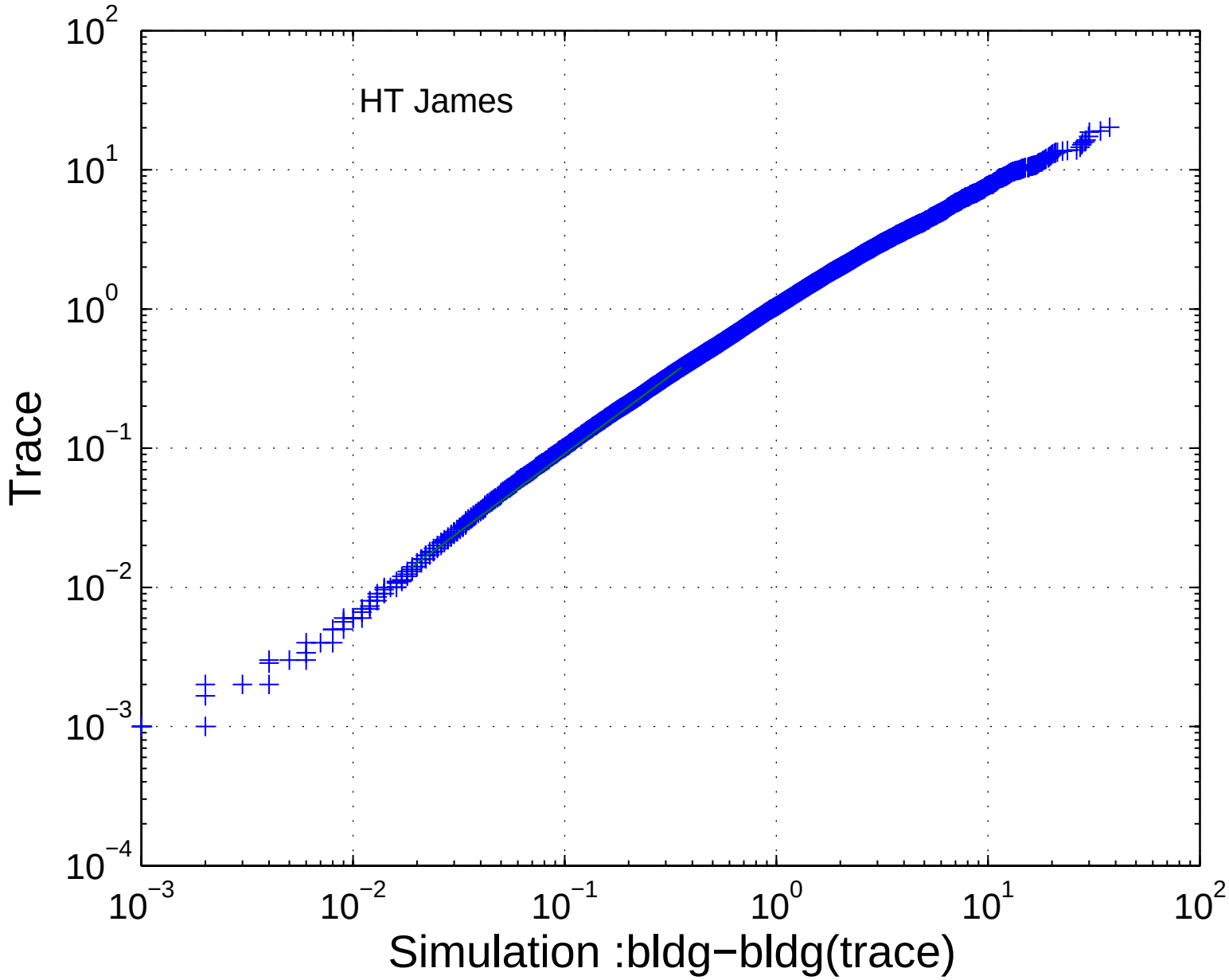
Hourly number of flow arrivals

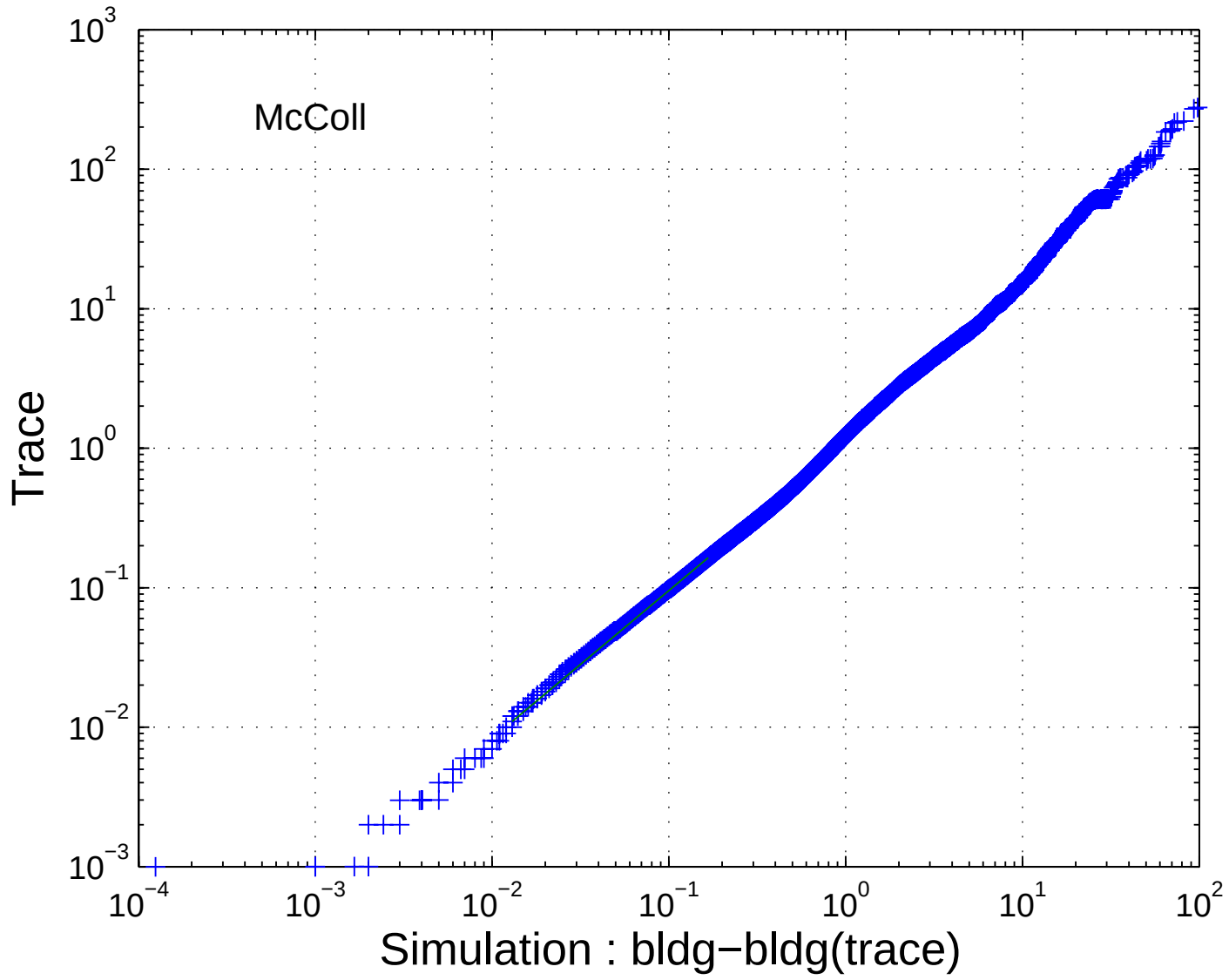


Autocorrelation of flow interarrivals




Davis −LIBRARY
Session arrivals
Time (hours)
0
10
20
30
40
50
60
70
80
90
0
20
40
60
80
100
120
140
160
180
200


HT James
Trace
Simulation :bldg−bldg(trace)
10^{2}
10^{1}
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}


McColl
Trace
Simulation : bldg−bldg(trace)
10^{3}
10^{2}
10^{1}
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}

Our models 2/2

Modeled variable	Model	Probability Density Function (PDF)	Parameters
Session arrival	Time-varying Poisson with rate	N: #sessions between t_1 and t_2 $\lambda = \int_{t1}^{t2} \lambda(t)dt, \Pr(N = n) = \frac{e^{-\lambda}\lambda^n}{n!}, n = 0,1,...$	Hourly rate: 44(min), 1132(max), 294(median)
AP of first association/session	Lognormal	$p(x) = \frac{1}{\sqrt{2\pi}x\sigma}\exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right]$	$\mu = 4.0855, \sigma = 1.4408$
Flow interarrival/session	Lognormal	Same as above	$\mu = -1.3674, \sigma = 2.785$
Flow number/session	BiPareto	$p(x) = k^{\beta}(1+c)^{\beta-\alpha}x^{-(\alpha+1)}(x+kc)^{\alpha-\beta-1}$ $(\beta x + \alpha kc), x \geq k$	$\alpha = 0.06, \beta = 1.72,$ $c = 284.79, k = 1$
Flow size	BiPareto	Same as above	$\alpha = 0.00, \beta = 0.91,$ $c = 5.20, k = 179$

Related work in wireless traffic modeling

- Over hourly intervals at AP-level
  - Captures finer spatial detail required for evaluating network functions with focus on AP-level (e.g., load-balancing, admission control)
  - Does not scale for large infrastructures
  - Data do not always amenable to statistical analysis
- Infrastructure-wide
  - Models amenable to statistical analysis
  - Concise summary of traffic demand at system-level
  - Fails to capture finer spatial detail required for evaluating network functions with focus on AP-level

Starting building & “roaming”

Small % of building-roaming flows

Little dependence on what kind of building a session is initiated