Understanding Wireless Network & Applications

Songwu Lu
UCLA Computer Science Department
slu@cs.ucla.edu

Goals

Understanding network & application behaviors using real traces

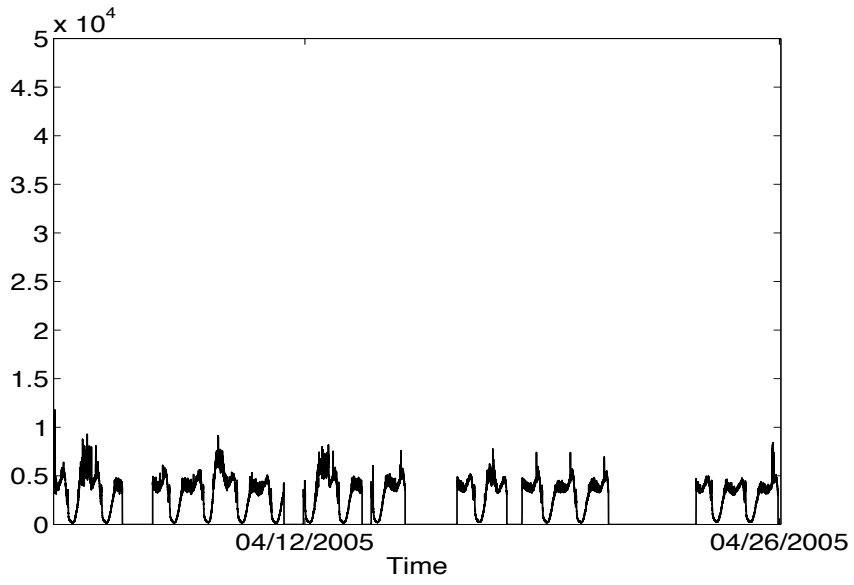
- average case
- pathological case

Two concrete topics:

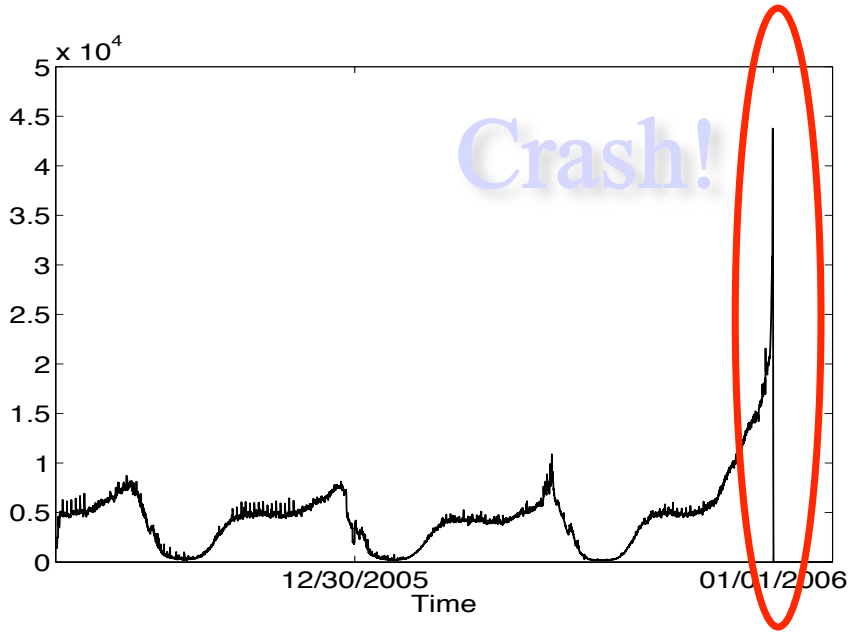
- Building application benchmarks over large WLANs
- Understanding SMS reliability over cellular networks

Normal & Pathological Cases

- Normal case
  - Occurring most time
  - Desired by operator and users
- Pathological case
  - Rarely occurring
  - Caused by dynamics
  - Undesired



SMS traffic in April 2005



SMS traffic in NYE 2005

Motivation

- Normal-case study
  - Average performance assessment → How well the network/service works in general?
    - Planning
    - Performance evaluation
- Pathological-case study
  - Robustness assessment → Whether network/service can sustain extreme conditions?
    - Robust design

Case 1: Normal Behaviors of TCP Flows over WLANs

- Application benchmarks
  - Average case behaviors
- Three issues:
  - Birth and Death process for a static flow
  - Location dependency of flows
  - Roaming flows

Traces

- Campus-wide
- 161 buildings
- 476 APs
- 3 months
- 2.17M flows

Primary trace

- Conference
- 1 subnet, 4 APs
- 3 days
- 299K flows

- CS building in a university
- 1 subnet,12 APs
- 12 weeks
- 74 hosts, 78M packets

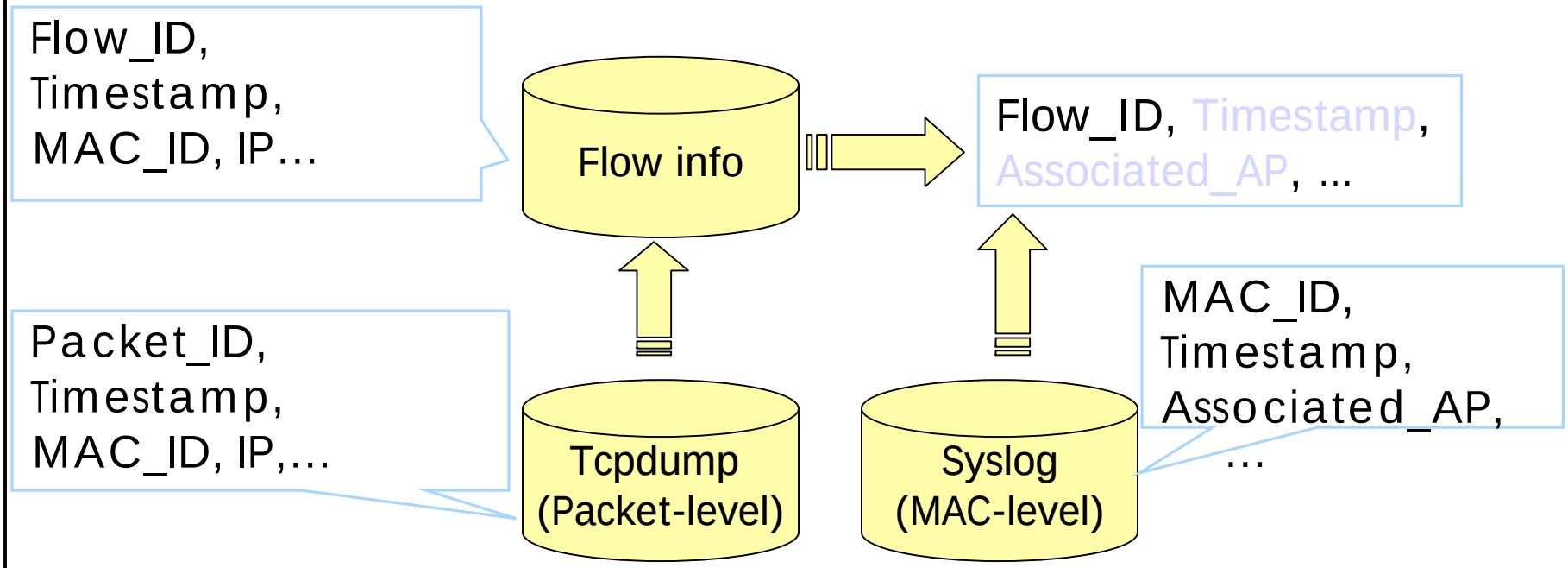
- Buildings in a company
- 177 APs
- 4 weeks
- 1366 users

Methodology

- Synthesize traces
- Analysis methods:
  - Two-level modeling: statistical models & regressions
  - Statistics: modeling matching etc.
  - Frequency analysis
- Cross-validation whenever possible
  - Check model validity using multiple datasets

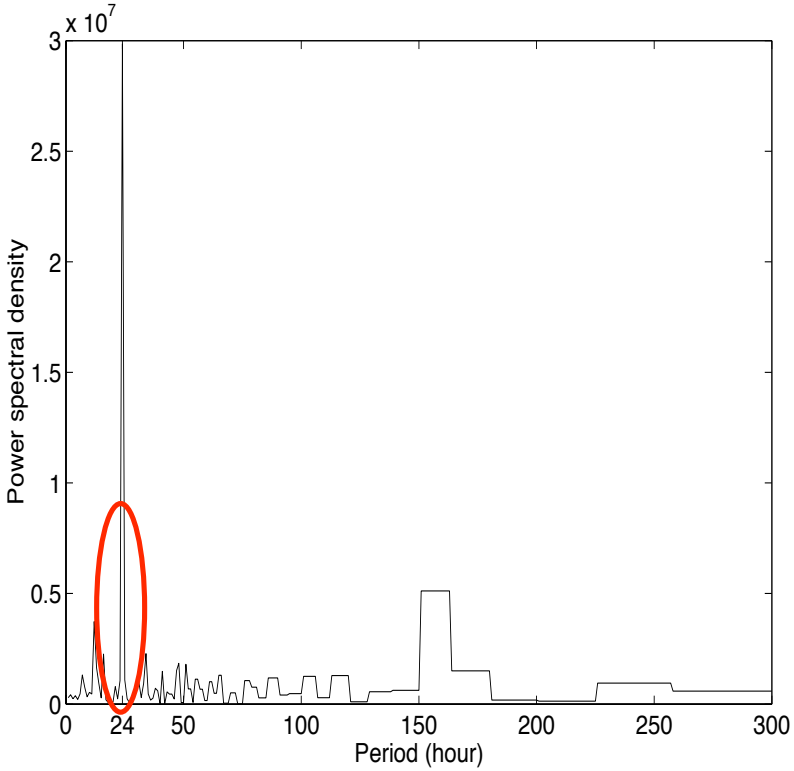
Mapping btw Ctrl & Data Planes

- To obtain both temporal and spatial information for flows

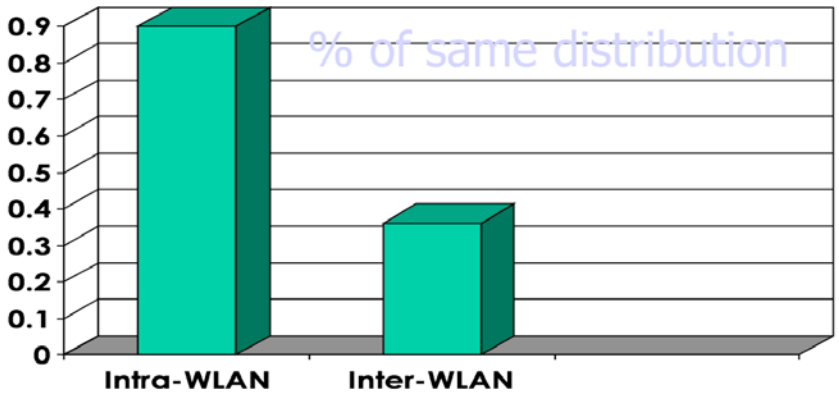


Static Flows

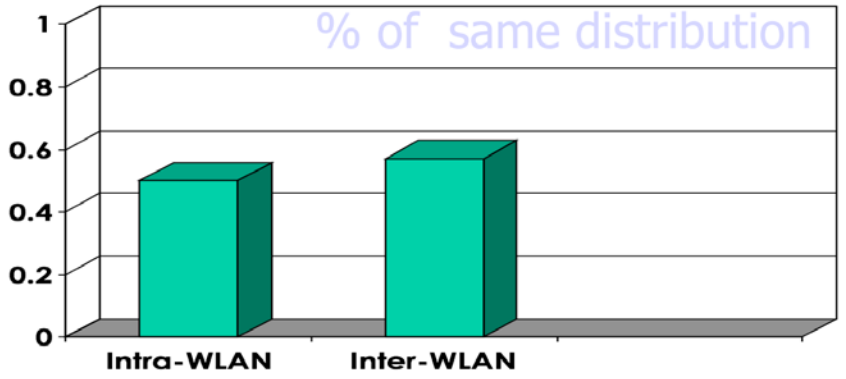
- Model inter-arrival time
  - Power spectral analysis
  - Match against statistical distributions
  - Regression
- Weibull regression model
  - Coexistence of long and short inter-arrival times, effect of time-of-day
- Flow size: Lognormal distribution
  - Consistent with Internet file size distribution



Spatial Correlation



Inter-arrival time



Flow size

- Implications:
  - APs in same WLAN observe similar flow arrival patterns
  - Flow size not relevant to where the flow comes from

Roaming Flows

- 0.13% flows are roaming while 60% hosts have mobility
  - Only 4.6% users generate traffic when they are physically moving
- Model roaming flows
  - Hand-off freq: match Geometry distribution

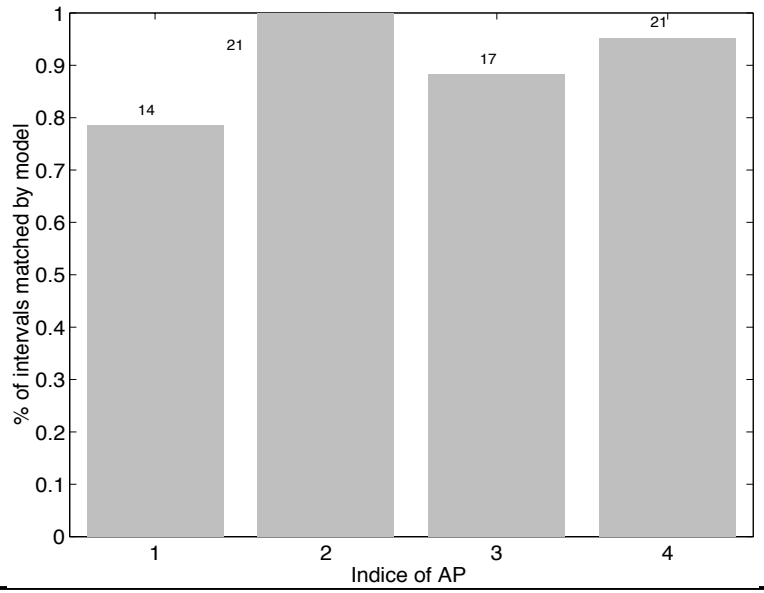
    Hand-off process is memoryless
  - Residing time: Weibull distribution

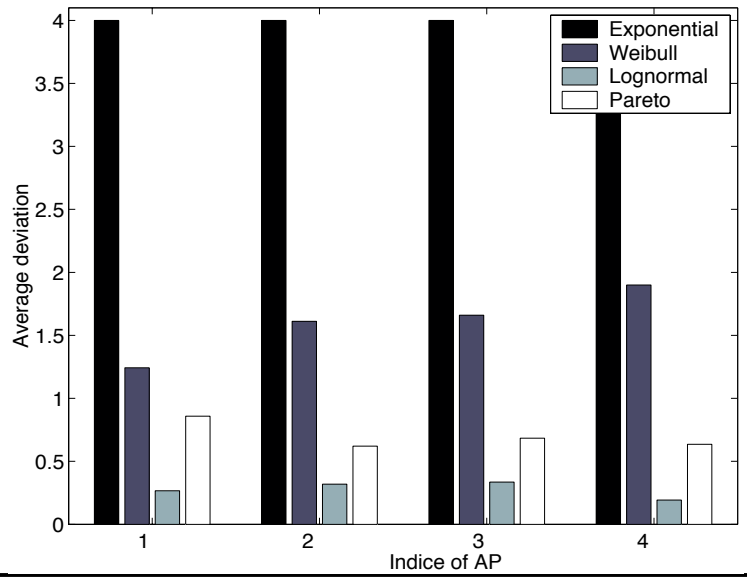
    Both short and long residing times exist

Cross Validation

Using UCSD01 dataset

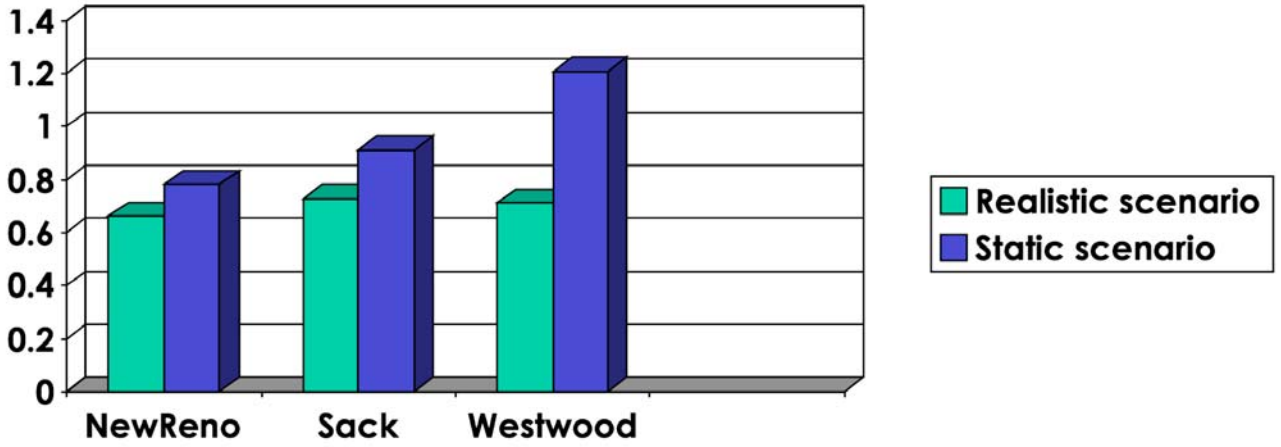
- Flow inter-arrival time: Weibull regression model matches among 78% time intervals
- Flow data size: Lognormal distribution gives the best match



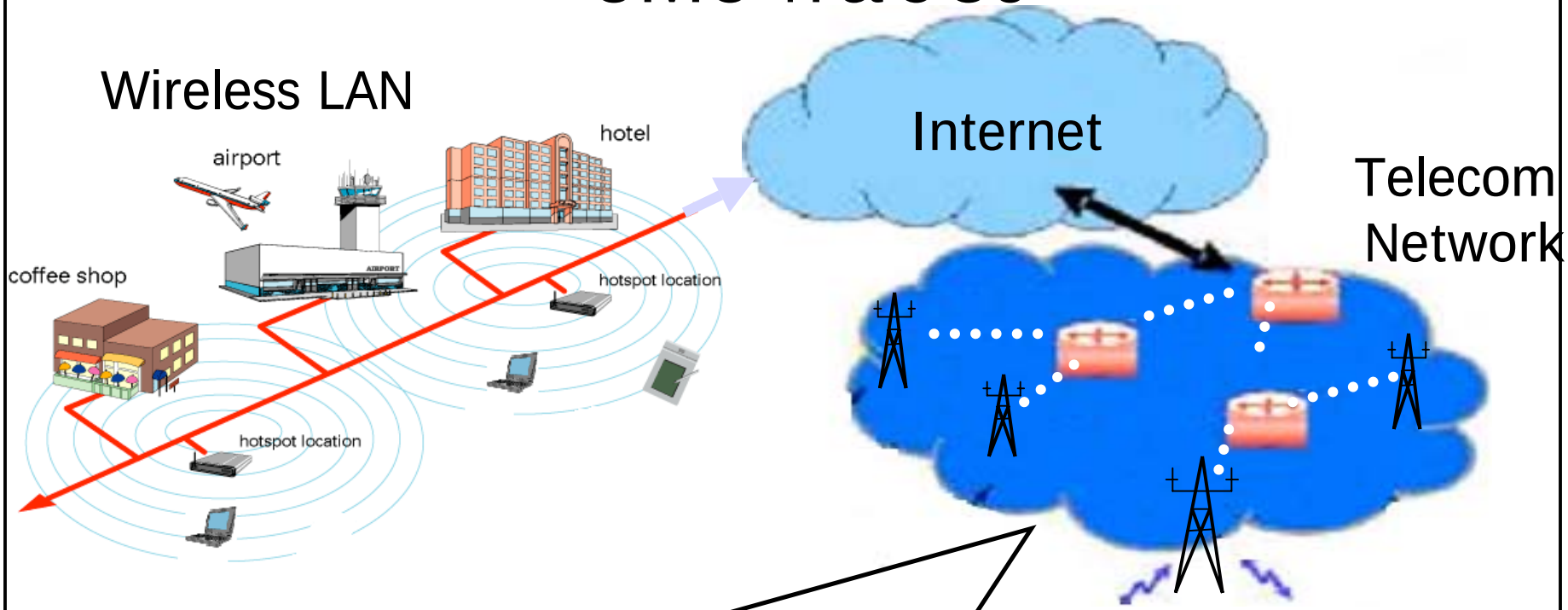


Applying Models in TCP

- Compare TCP throughput between
  - Static scenario used by literature
  - Realistic scenario: dynamic flow arrivals by using derived models
- 40% difference

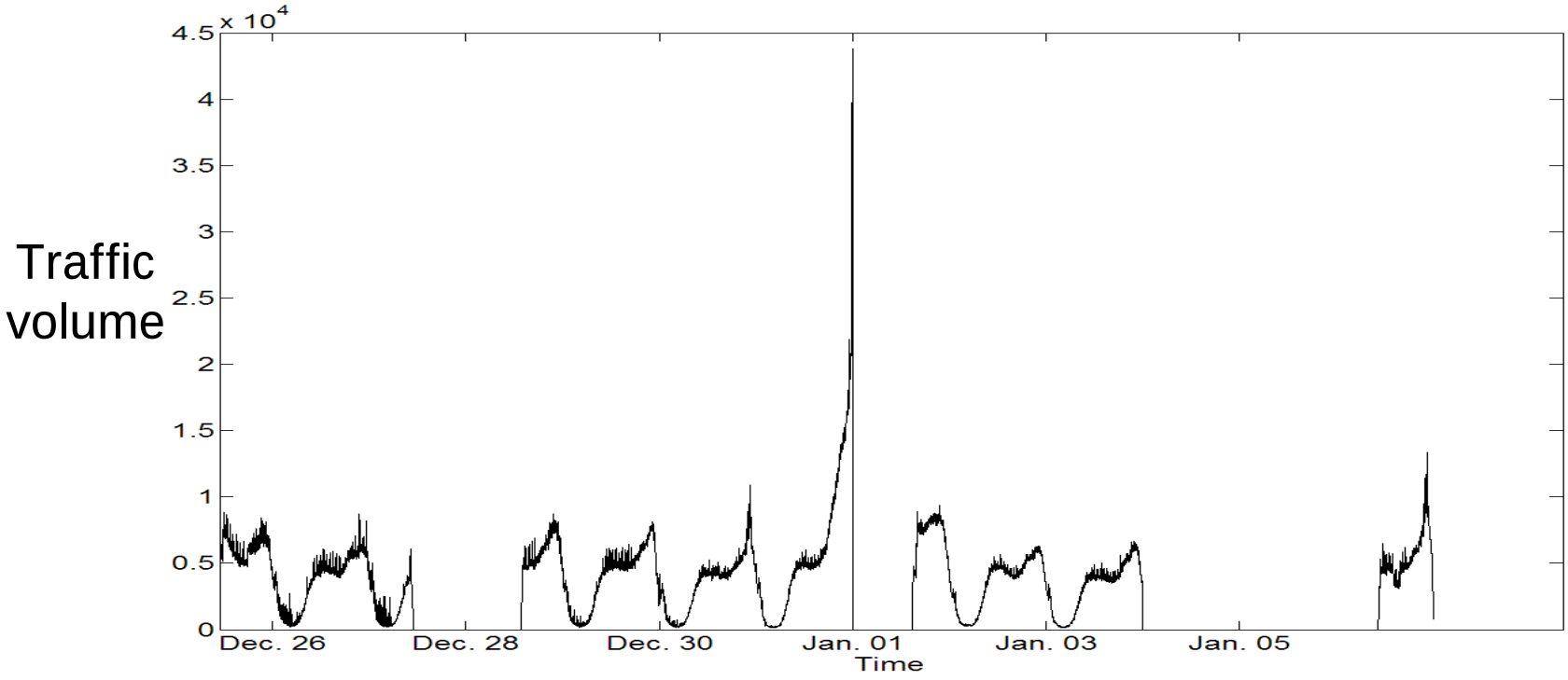


SMS Traces



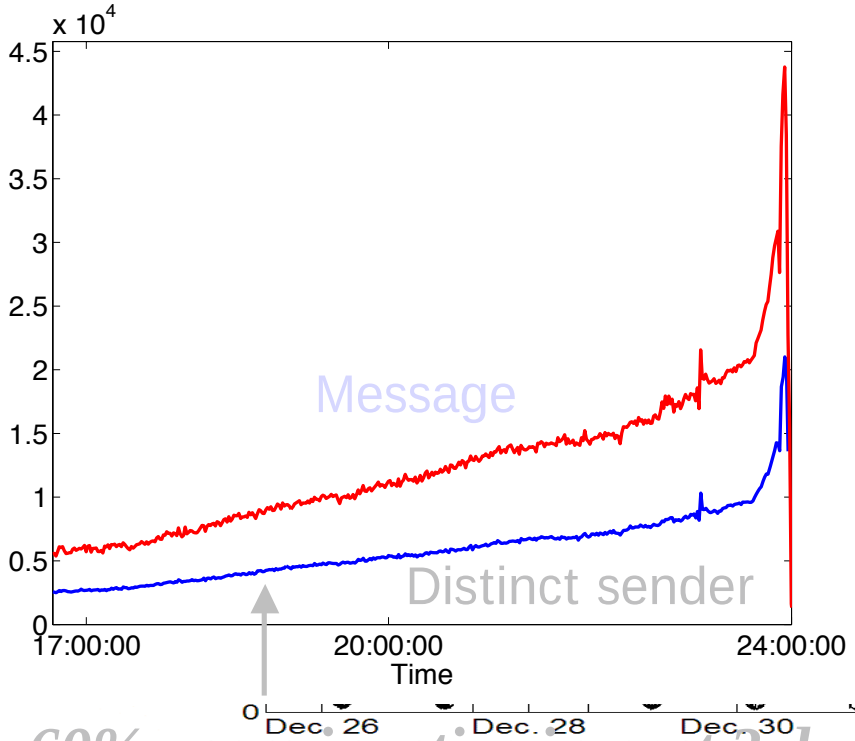
- A large national carrier with 10-million users
- 59-million messages in traces
- 4 weeks of measurements in 2004- 2005

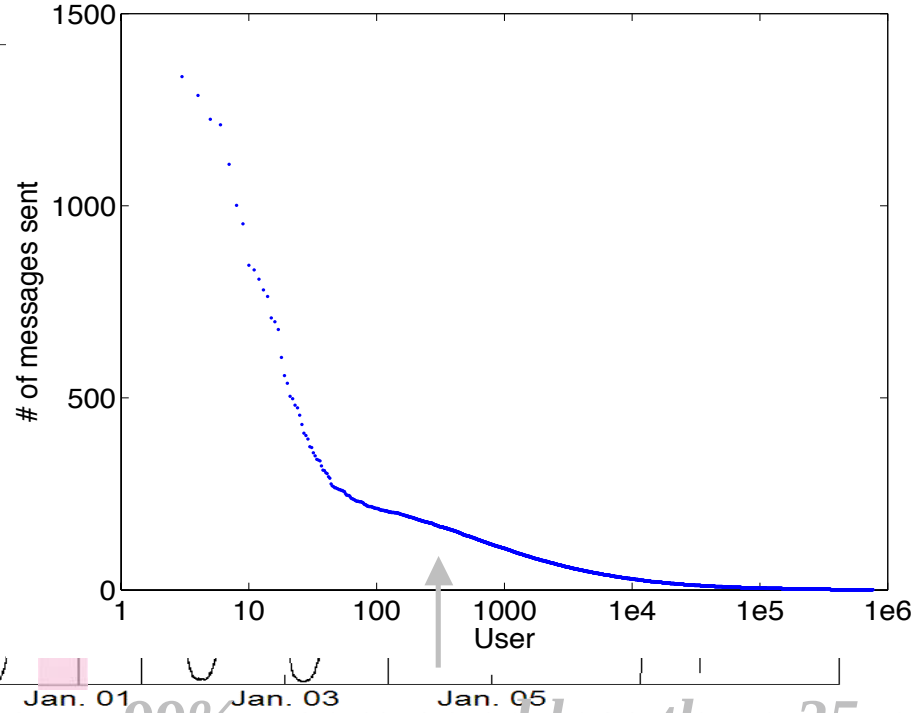
SMS Crash in New Year Eve



- Approach: mining traces to infer underlying user behavior
  - Providing suggestions for provisioning

Analyzing User Behavior





60% users inactive in past 3 days *99% users send less than 35 msgs*

- Light-tailed distribution of messages per user
  - Conjecture: billing policy
- Flash event is caused by a sudden resurrection of many users in a short time

Learning from Our Experiences

- Challenges faced
  - The good
  - The bad
  - The ugly
- What more needed

Insights

- Cross Validation among traces is important !
  - Applicability of results
  - Problems: Quality of data traces varies a lot
    - Darthmouth is the best, UCSD is OK
    - Stanford and IBM are not good
- Synthesizing data traces
  - Synchronization issue

The Good, The Bad, The Ugly

- The good:
  - Lots of traces available now
  - More tools are coming up
- The bad:
  - hard to be deep, easy to be superficial
  - Data are missing when needed most
    - Cases of failure, attacks, extreme overload
- The ugly:
  - Trace processing is time consuming
  - Dig gold out of dirt: Need an agenda first

What more needed

- better trace/data analysis methods and tools
  - Spatial-temporal dynamics in wireless
  - Need interdisciplinary effort
    - Statistics, signal processing, data mining, …
- Visualization tools
  - Visualize changes in both time and space
  - Real-time monitoring
- Meaningful Benchmarks
  - Capture both normal behaviors and extreme conditions