ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

Ασύρματα Δίκτυα και Κινητοί Υπολογισμοί

Μαρία Παπαδοπούλη

Τμήμα Επιστήμης Υπολογιστών

Πανεπιστήμιο Κρήτης

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Κρήτης**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Σημείωμα αδειοδότησης

Empirical measurements

- Can be beneficial in revealing
  - deficiencies of a wireless technology
  - different phenomena of the wireless access & workload
- Provide data for modelling efforts aiming to produce **more realistic models & synthetic traces**
- Enable **meaningful performance analysis studies** using such empirical, synthetic traces and models

- **Network performance benchmarks** include:
- Jitter
- Latency/delay/response delay
- One way, round-trip delay

Network workload can be characterized based on:

- Amount of traffic
- Number of packets or bytes
- Downloading vs. uploading
- Packet or flow arrival process
- Interactivity model
- Application type
- Usage pattern

Network topologies can be described based on

- Their connectivity
- Link charactertsics
- Distribution & density of peers
- Degree of clustering
- Co-residency time
- Inter-contact time
- Duration of disconnection for the Internet
- Interaction patterns

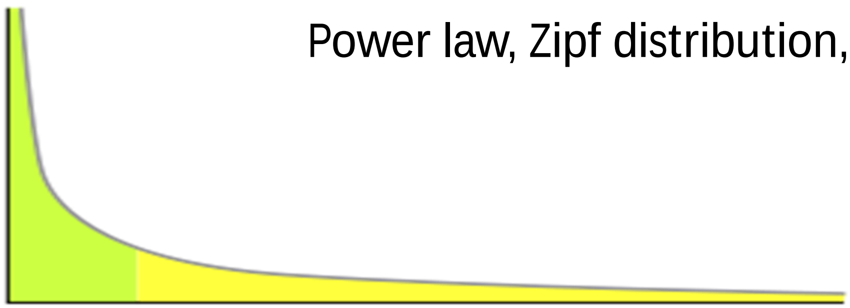
Internet traffic (amount of bytes) exhibits a self-similar nature.
Web traffic also exhibits self-similarity.
The majority of web traffic in wired networks is below 10KB, while a small percentage of vey large flow account for 90% of the total traffic. Power laws can describe web flow sizes.

So Poisson processes cannot accurately model the traffic load. However, they can be used to model the arrival of user sessions (e.g., telnet connections or arrivals at a wireless AP).

Poisson process

- Stochastic process which counts the number of events and the time that these events occur in a given time interval.
- The time between each pair of consecutive events has an **exponential distribution** with parameter λ and each of these inter-arrival times is assumed to be independent of other inter-arrival times

Exponential distribution : PDF X: Prob(X=x)=$\lambda e^{-\lambda x}$, $x \in [0, \infty)$, $E[X] = 1/\lambda$, $Var[X] = 1/\lambda^2$

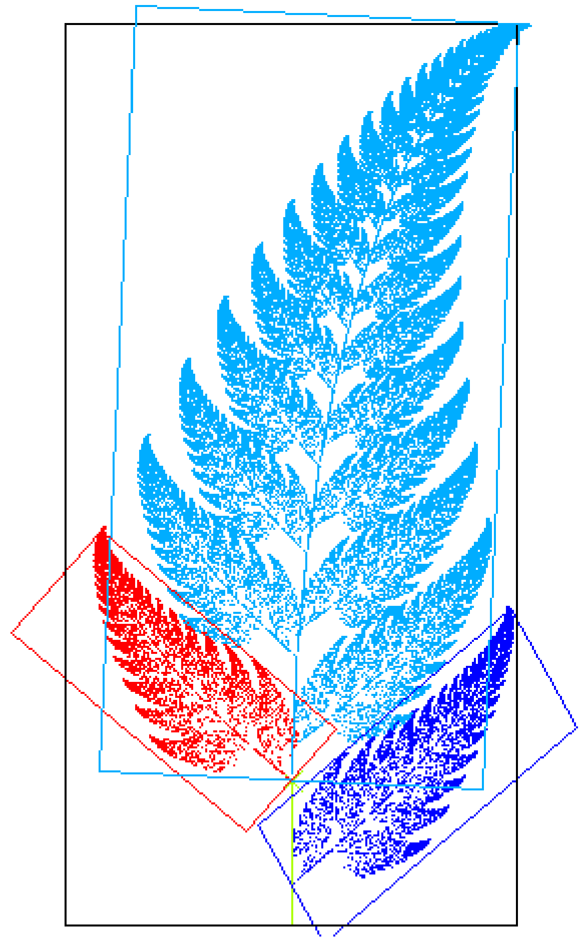


Power law, Zipf distribution, Pareto distribution are "heavy-tail" distributions

Pareto: $$\overline{F}(x) = \Pr(X > x) = \begin{cases} \left(\frac{x_m}{x}\right)^{\alpha} & \text{for } x \geq x_m, \\ 1 & \text{for } x < x_m. \end{cases}$$

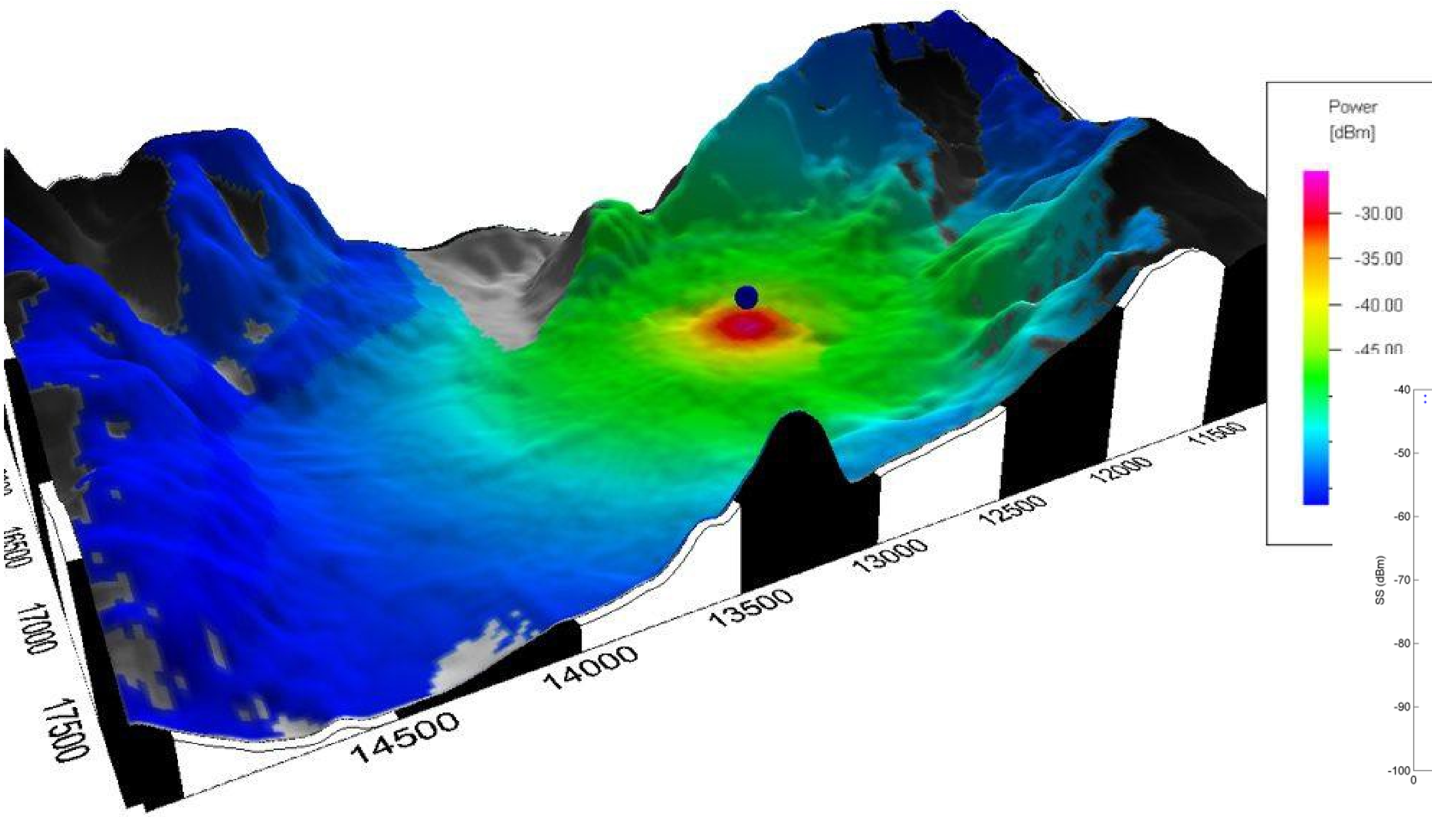
a **self-similar** object is exactly or approximately similar to a part of itself (i.e. the whole has the same shape as one or more of the parts).

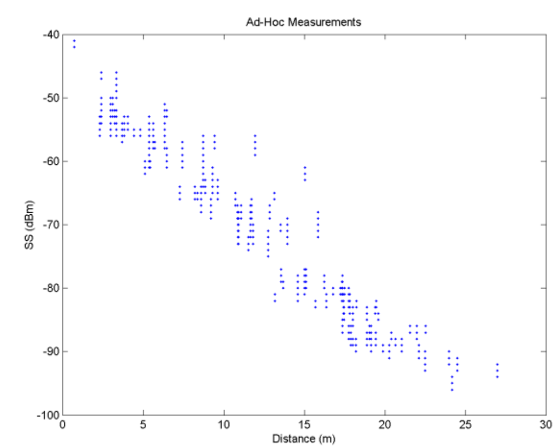
A self-similar object "looks" roughly the same on any scale



Propagation Models

- One of the ***most difficult part of the radio channel design***
- Done in **statistical fashion** based on measurements made specifically for an intended communication system or spectrum allocation
- Predicting the **average signal strength** at a given **distance from the transmitter**





Signal Power Decay with Distance

- A signal traveling from one node to another experiences fast (multipath) fading, shadowing & path loss
- ***Ideally,*** ***averaging RSS*** over sufficiently **long time interval**

excludes ⇨ the effects of multipath fading & shadowing

general ***path-loss model***:

$$\bar{P}(d) = P_0 - 10n \log_{10} (d/d_o)$$

n: path loss exponent

P(d): the average received power in dB at distance ***d*** P_0 is the received power in dB at a short distance d_0

Signal Power Decay with Distance

- ***In practice***, the observation interval is not **long enough** to mitigate the effects of shadowing

☞ The received power is commonly modeled to include both **path-loss** & ***shadowing*** effects, the latter of which are modeled as a zero-mean Gaussian random variable with standard deviation σ_{sh} in the logarithmic scale, P(d), in dB can b_e expressed:

$P(d) \sim N(P(d), \sigma_{sh}^2)$

This model can be used in both line-of-sight (LOS) & NLOS scenarios with appropriate choice of channel parameters

Important aspects of monitoring

- Identification of the **dominant parameters**
- based on the aspects you need to measure, the parameters that need to be monitored are decided ...
- **Strategic placement of monitors**
- e.g., at routers, APs, clients, and other devices, ...
- **Automation of the monitoring** process to reduce human intervention in managing the monitors and collecting data
- **Aggregation of data collected** from distributed monitors to improve the accuracy, while maintaining a low communication and energy overhead
- (cross-)validation study to verify that the collected traces correspond to representative conditions

Monitoring

- Depending on type of conditions that need to be measured, monitoring needs to be performed at
    - Certain layers
    - **Spatio-temporal** granularities
- Monitoring tools
  - Are **not** without flaws
  - Several issues arise when they are used in parallel for thousands devices of different types & manufacturers:
    - ***Fine-grain*** **data sampling**
    - **Time *synchronization***
    - ***Incomplete*** **information** (missing values, incorrect values)
    - Data ***consistency***
    - ***Vendor-specific*** information & dependencies (often not publicly available)

Monitoring tools

- *Fine-grain* data sampling

What is its spatio-temporal granularity?

- Time *synchronization*

Clocks have different drifts…

- *Incomplete* information (missing values, incorrect values)
    - How do you handle missing values? Explain what caused them!
    - Various techniques to "fill" the gaps, if necessary, such as interpolation, model-prediction, matrix completion and other sophisticated statistical analysis methods …
    - Explain the outliers! Are there due to misconfigured devices? Or network anomalies? Or due to the occurrence of "extreme" phenomena?
- Data *consistency*
- *Vendor-specific* information & dependencies (often not publicly available)

    e.g., RSSI differs depending on the manufacture

Monitoring & Data Collection

- Fine spatio-temporal detail monitoring can

👍 Improve the **accuracy** of the performance estimates but also

👎 Increase the **energy spendings** and **detection delay**

Network interfaces may need to

- Monitor the channel in **finer & longer time scales**
- Exchange this information with other devices

Challenges in Monitoring (1/2)

- **Identification** of the dominant parameters through
  - ***sensitivity analysis*** studies
- **Strategic placement** of monitors at
  - Routers
  - APs
  - clients and other devices of users
- **Automation** of the monitoring process to *reduce human intervention* in managing the
    - Monitors
    - Collecting data

Challenges in Monitoring (2/2)

- Aggregation of data collected from distributed monitors to improve the accuracy while maintaining low overhead in terms of
  - Communication
  - Energy

- ***<u>Cross-layer</u>*** *measurements*, collected data spanning from the physical layer up to the application layer, are required

Wireless Networks

– Are extremely complex
– Have been used for many different purposes
– Have their own distinct characteristics due to radio propagation characteristics & mobility

e.g., wireless channels can be ***highly asymmetric*** and ***time varying***

Note:

Interaction of different layers & technologies creates many situations that cannot be foreseen during design & testing stages of technology development

Empirically-based Measurements

- Real-life measurement studies can be particularly beneficial in revealing
  - deficiencies of a wireless technology
  - different phenomena of the wireless access and the workload
- Rich sets of data can
  - Impel modeling efforts to produce more realistic models
  - Enable more meaningful performance analysis studies

It is important to assess whether typical assumptions are realistic!

Typical assumptions in performance analysis of wireless networks

- Models & analysis of wired networks are valid for wireless networks
- Wireless links are ***symmetric***
- Link conditions are ***static***
- The density of devices in an area is ***uniform***
- The communication pairs are ***fixed***
- User mobility is based on ***random-walk models***

Wireless Access Parameters

- Traffic workload
  - In different time-scales
  - In different spatial scales (e.g., AP, client, infrastructure)
  - In bytes, number of packets, number of flows, application-mix
- Delays
  - Jitter and delay per flow
  - Statistics at an AP and/or channel
- Packet losses

  %packet loss & burstiness of packet losses

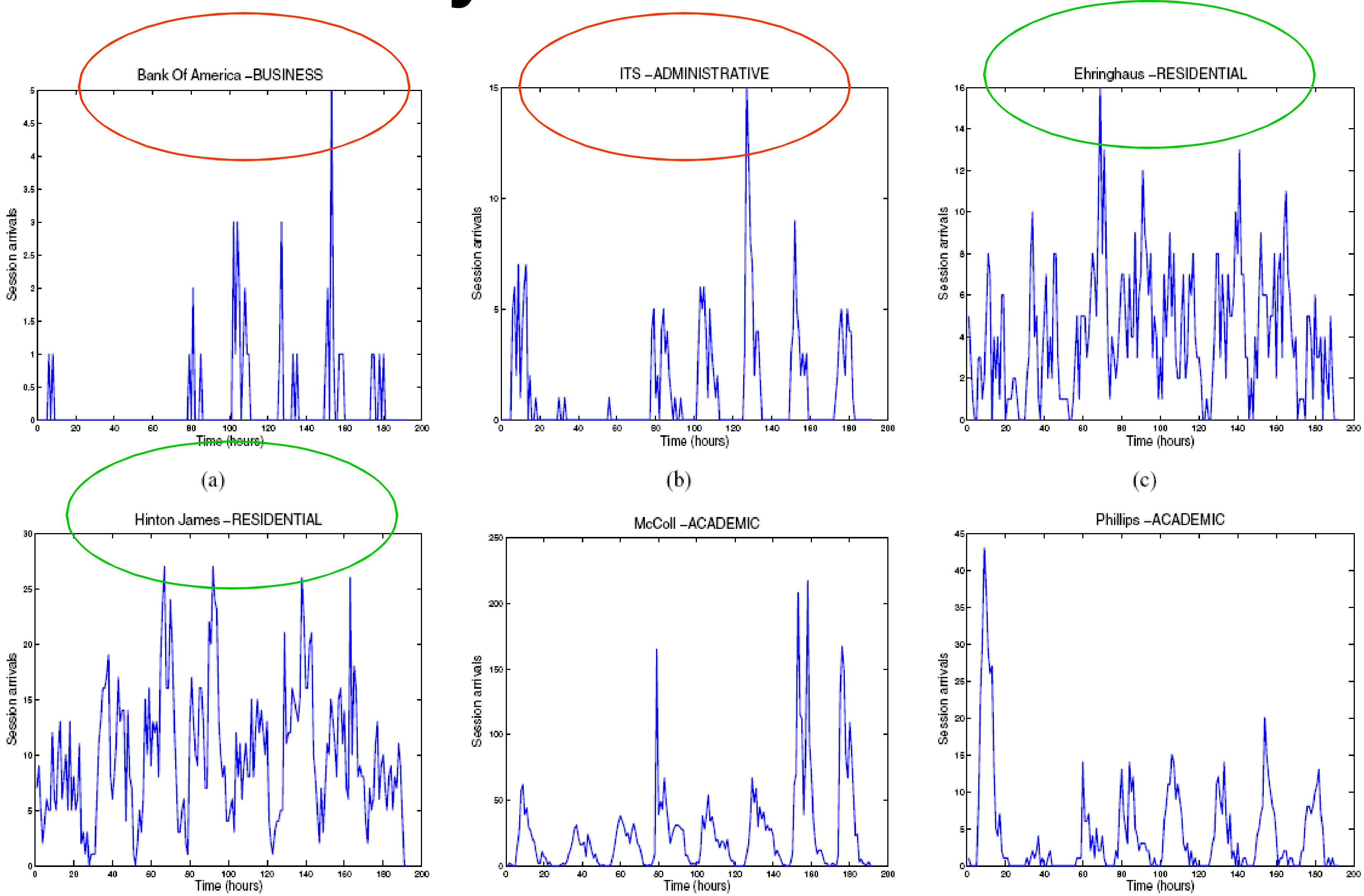
- User mobility patterns
- Link conditions & channel quality
- Network topology

Traffic Load Analysis

- As the wireless user population increases, the characterization of traffic workload can facilitate
  - More efficient network management
  - Better utilization of users’ scarce resources

- Application-based traffic characterization

Hourly Session arrival rates



Traffic Load at APs

- Wide range of workloads that log-normality is prevalent
- In general, traffic load is light, despite the long tails
- No clear dependency with type of building the AP is located exists
  - Although some stochastic ordering is present in
    - Tail of the distributions
- Dichotomy among APs is prominent in both infrastructures:

  APs dominated by uploaders

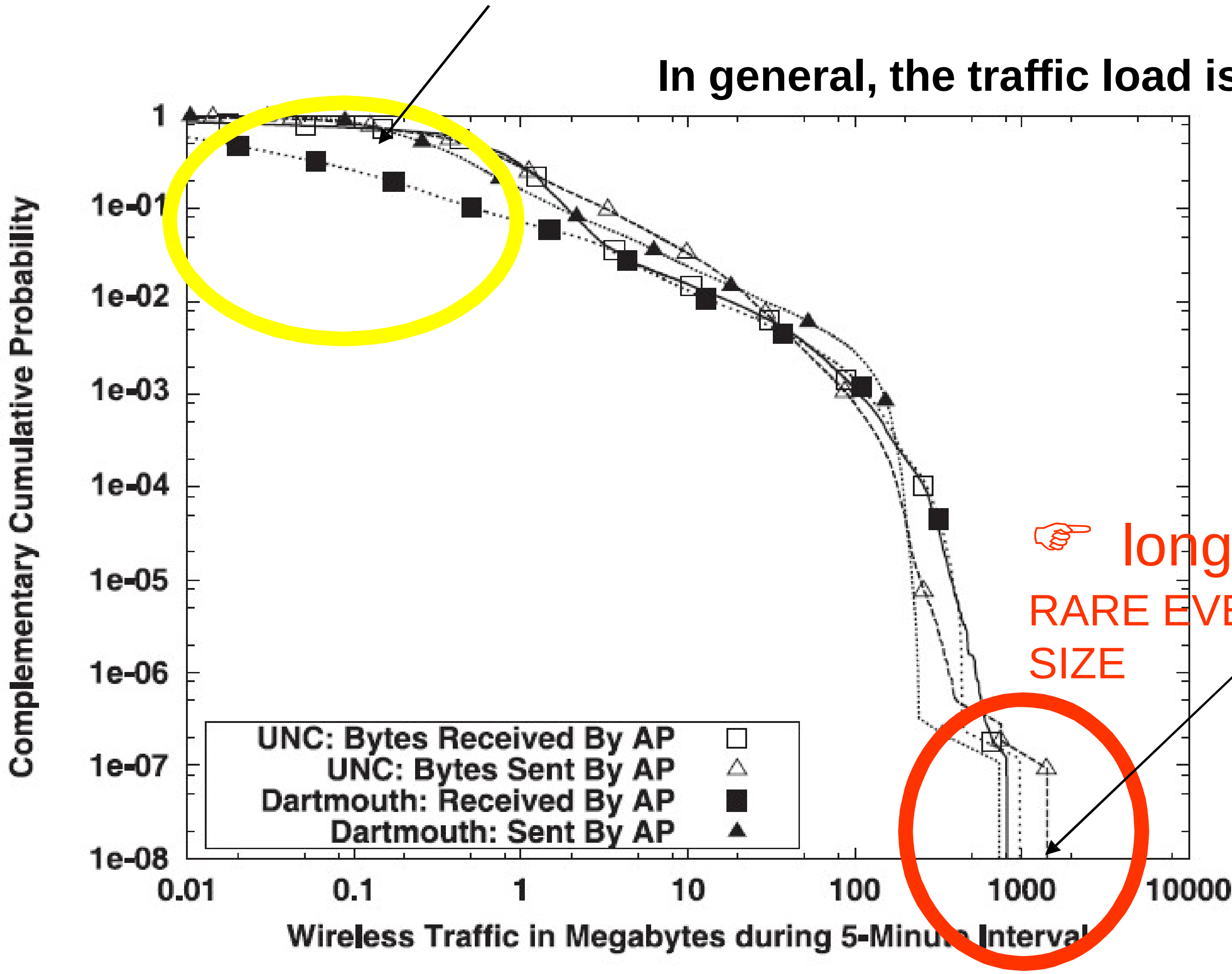
  APs dominated by downloaders

- As the total received traffic at an AP ↑
  - There is also ↑ in its total traffic sent
  - ↓ in the sent-to-received ratio

Traffic load at APs

- **Substantial number of non-unicast** packets
- Number of unicast received packets strongly correlated with number of unicast sent packets (in log-log scale)
- Most of APs send & receive packets of relatively **small size**
- Significant number of APs show rather **asymmetric packet sizes**
  - APs with **large sent** & **small receive packets**
  - APs with **small sent & large receive packets**
- Distributions of the number of associations & roaming operations are heavy-tailed
- **Correlation between the traffic load & number of associations** in log-log scale

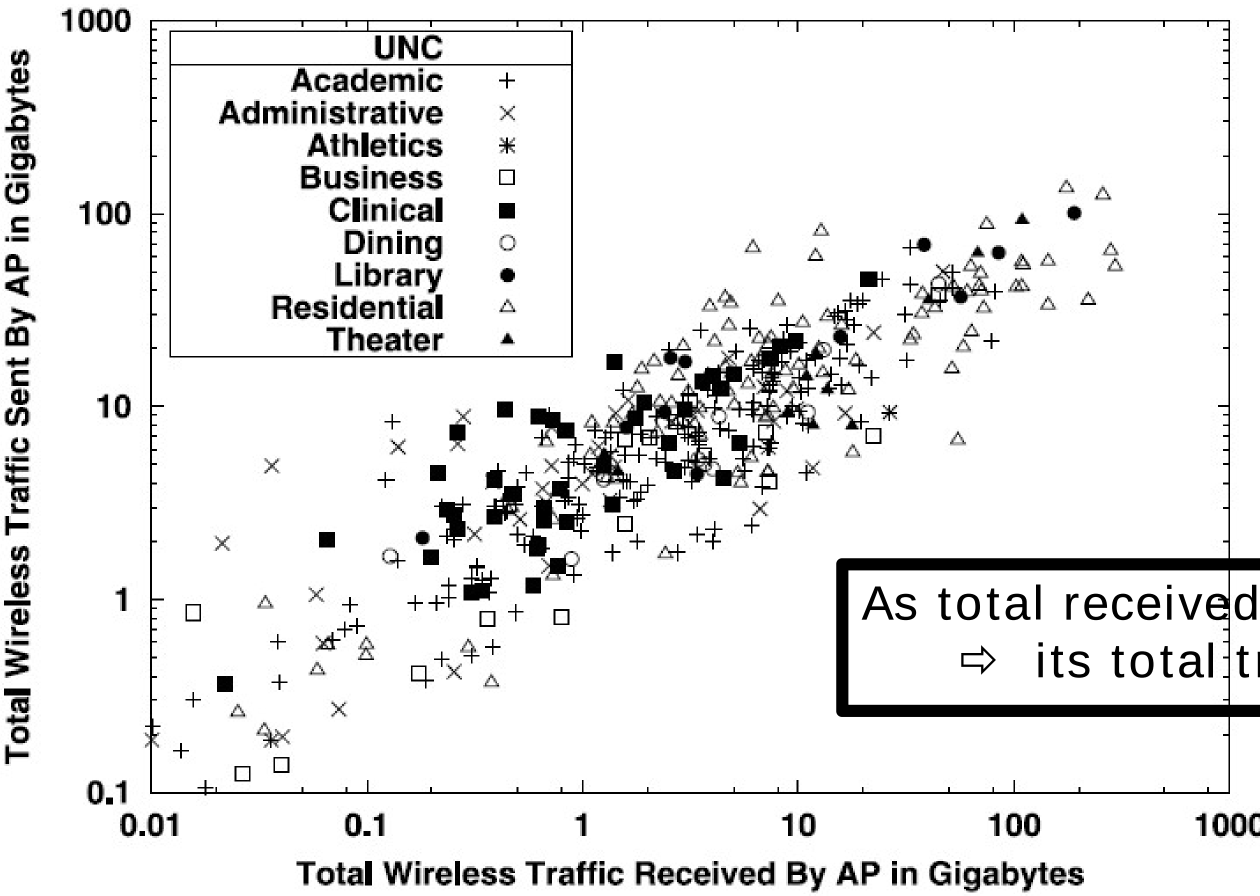
In general, the traffic load is light



☞ long tails

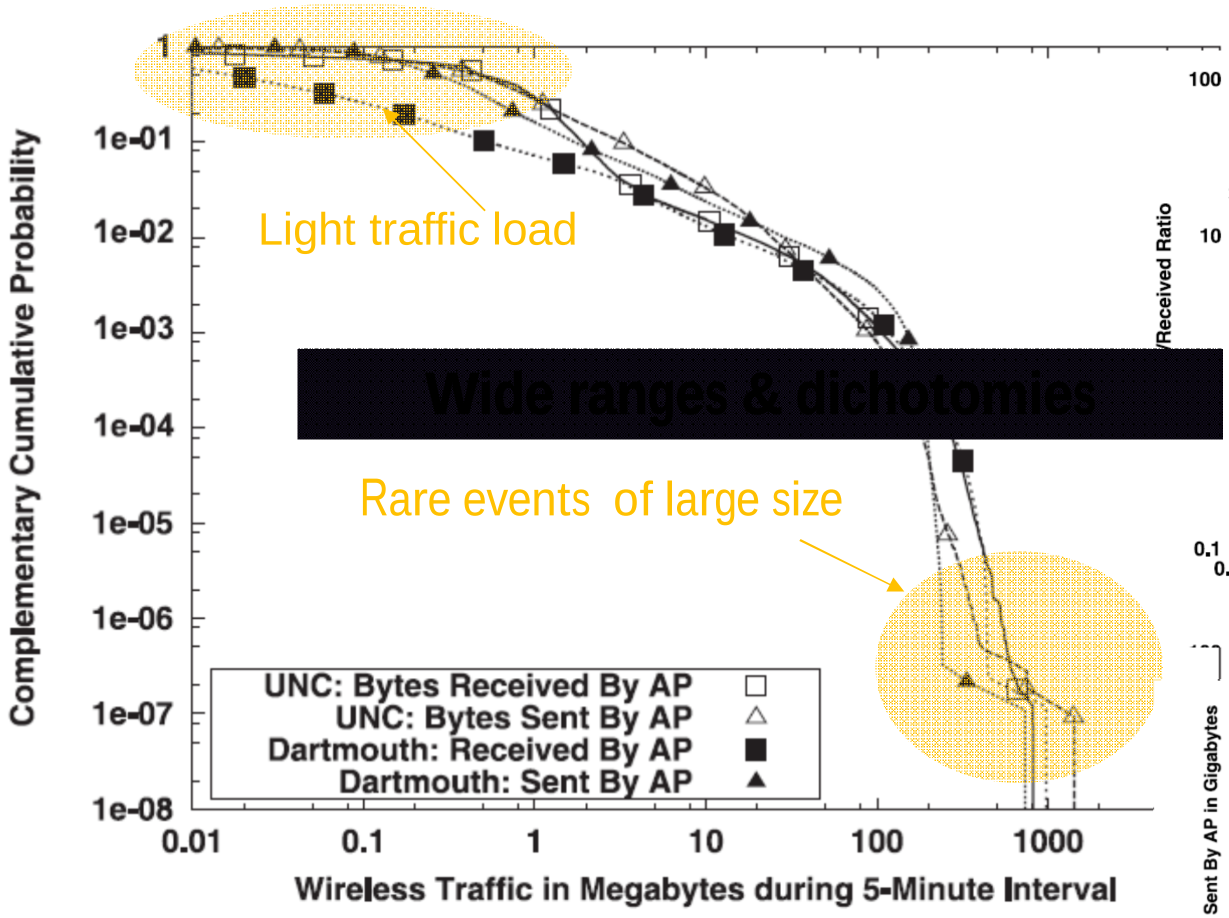
RARE EVENTS OF LARGE SIZE

Wide-range of workloads

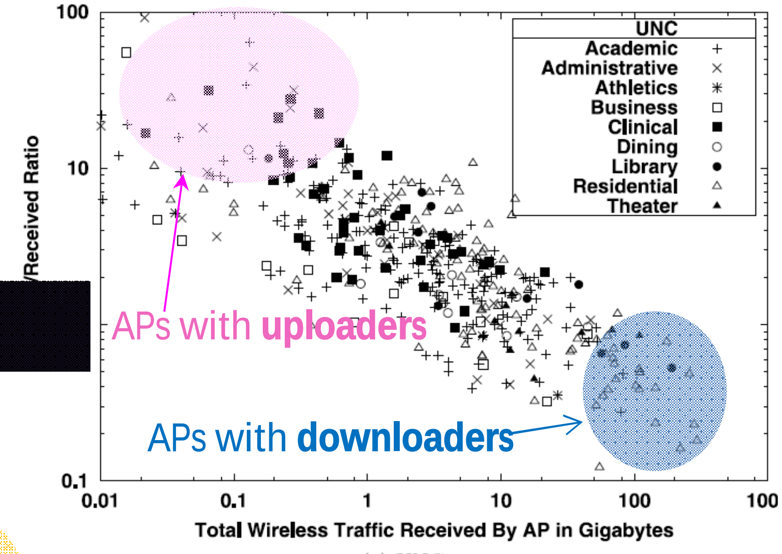


(b) UNC

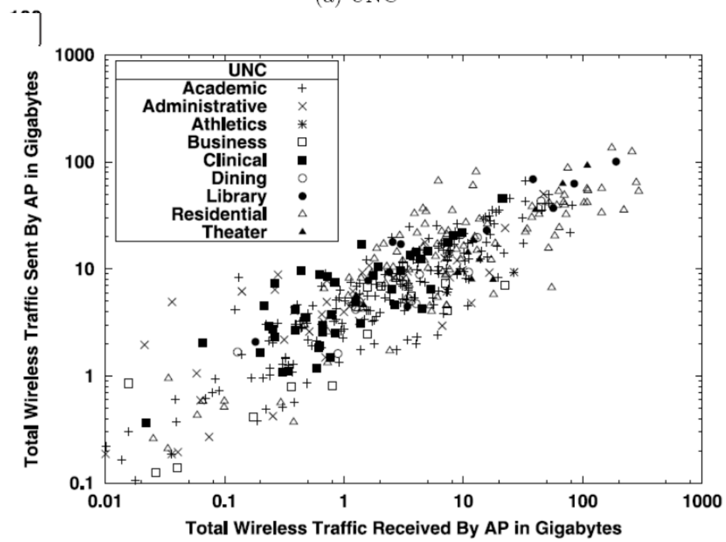
As total received traffic at an AP ↑
⇨ its total traffic sent ↑



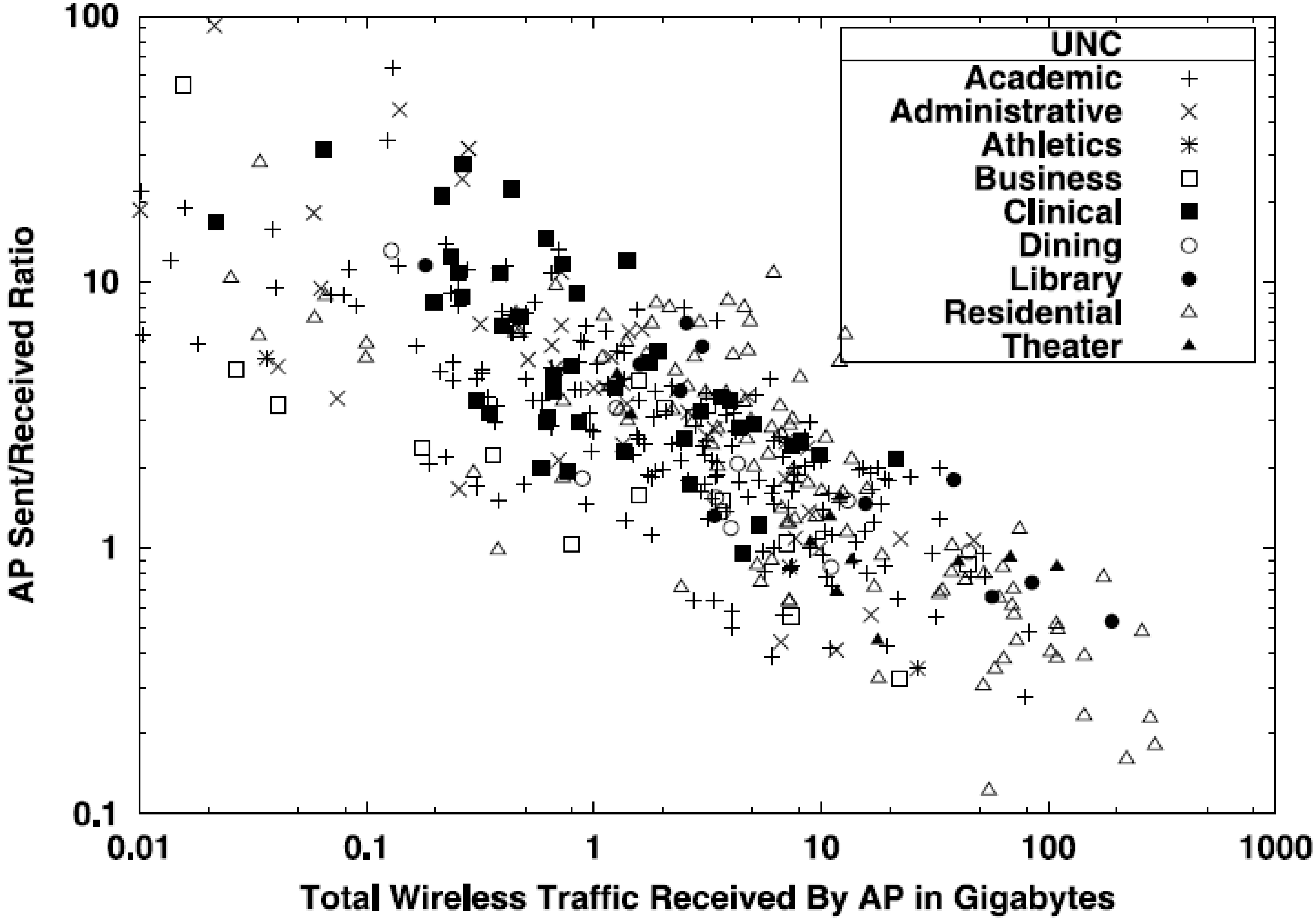
[redacted]



(a) UNC



(b) UNC


UNC
Academic
Administrative
Athletics
Business
Clinical
Dining
Library
Residential
Theater
100
10
1
0.1
0.01
0.1
1
10
100
1000
AP Sent/Received Ratio
Total Wireless Traffic Received By AP in Gigabytes

(a) UNC

Application-based Traffic Characterization

👎 Using **port numbers to classify flows** may lead to significant amounts of **misclassified traffic** due to:

–**Dynamic** port usage

–**Overlapping** port ranges

–Traffic masquerading

•Often peer-to-peer & streaming applications:

- Use dynamic ports to communicate
- Port ranges of different applications may overlap
- May try to masquerade their traffic under well-known “non-suspicious” ports, such as port 80

Desirable Properties for Models

- Accuracy
- Tractability
- Scalability
- Reusability
- “Easy” interpretation

Related work

- Rich literature in **traffic characterization in wired networks**
  - *Willinger, Taqqu, Leland, Park* on **self-similarity of Ethernet LAN traffic**
  - *Crovela, Barford* on **Web traffic**
  - Feldmann, Paxson on **TCP**
  - *Paxson, Floyd* on **WAN**
  - *Jeffay, Hernandez-Campos, Smith* on **HTTP**

•
•
•

- **Traffic generators for wired traffic**
  - Hernandez-Campos, Vahdat, Barford, Ammar, Pescape, ...
- **P2P traffic**
  - Saroiu, Sen, Gummadi, He, Leibowitz, ...
- **On-line games**
  - Pescape, Zander, Lang, Chen, ...
- **Modelling of wireless traffic**
  - Meng et al.

Dimensions in Modeling Wireless Access

- Intended user demand
- User mobility patterns
  - Arrival at APs
  - Roaming across APs
- Channel conditions
- Network topology

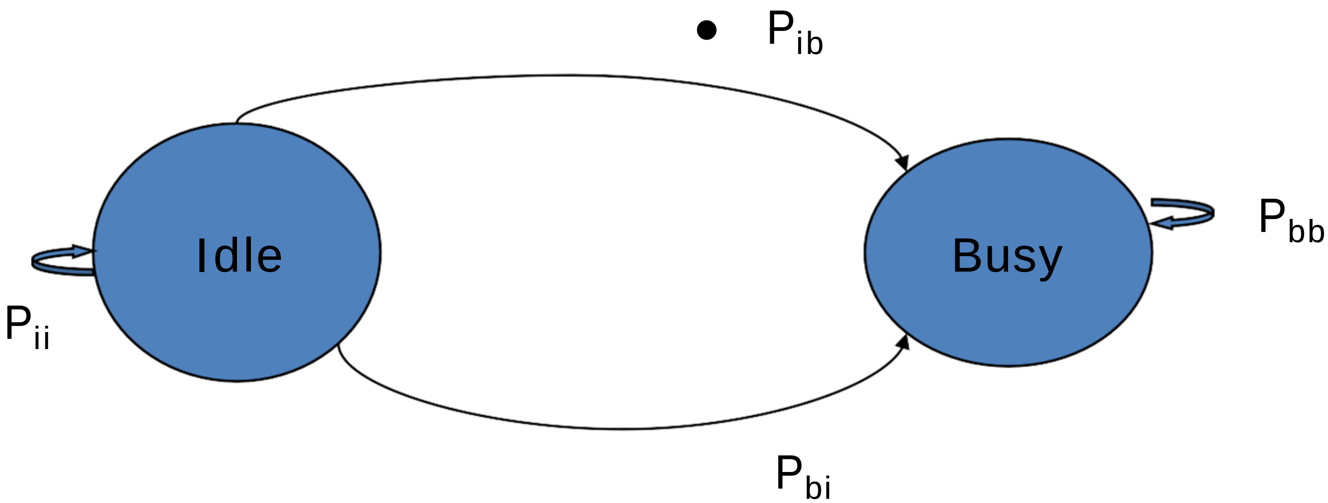
Mobility models

- Group or individual mobility
- Spontaneous or controlled
- Pedestrian or vehicular
- Known a priori or dynamic

- Random-walk based models
  - Randway model in ns-2
- Markov-based model

A Very Simple Channel Model

- **Gilbert model**
- ☺ Compute ***stationary probabilities***



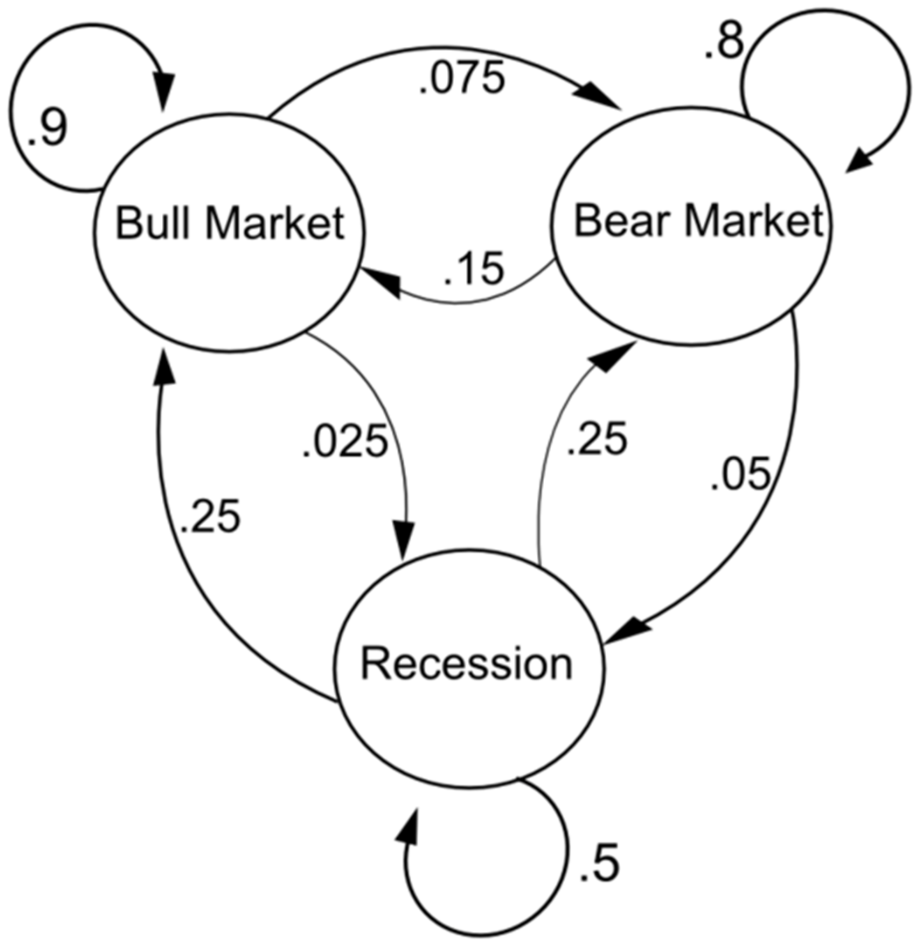
- A channel can be in the idle or busy state
- Markov-based model allows us to determine:
  - How much time the system spends in each state
  - Probability of being in a particular state

☞ In real rife, there is non-stationarity due ***dynamic phenomena***

Markov chain model:
(definition)
random process usually characterized as memoryless: the next state depends only on the current state and not on the sequence of events that preceded it.

Example:

The states represent whether the economy is in a bull market, a bear market, or a recession, during a given week:
a bull week is followed by another bull week 90% of the time, a bear market 7.5% of the time, and a recession the other 2.5%.



Label the state space
{1=bull, 2=bear, 3=recession}

Define the **transition matrix**:

$$P = \begin{bmatrix} 0.9 & 0.075 & 0.025 \\ 0.15 & 0.8 & 0.05 \\ 0.25 & 0.25 & 0.5 \end{bmatrix}.$$

(Continuing from the previous slide)

From this figure, it is possible to calculate the **long-term fraction of time** during which the economy is in a recession, or on average, how long it will take to go from a recession to a bull market

The distribution over states can be written as a stochastic row vector *x* with the relation $x^{(n+1)} = x^{(n)} P$.

So if at *time n* the system is in the state 2 ("bear") then, at *time n + 3* the distribution is

$$
\begin{aligned}
x^{(n+3)} &= x^{(n+2)} P = \left(x^{(n+1)} P\right) P \\
&= x^{(n+1)} P^2 = \left(x^{(n)} P^2\right) P \\
&= x^{(n)} P^3 \\
&= \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0.9 & 0.075 & 0.025 \\ 0.15 & 0.8 & 0.05 \\ 0.25 & 0.25 & 0.5 \end{bmatrix}^3 \\
&= \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0.7745 & 0.17875 & 0.04675 \\ 0.3575 & 0.56825 & 0.07425 \\ 0.4675 & 0.37125 & 0.16125 \end{bmatrix} \\
&= \begin{bmatrix} 0.3575 & 0.56825 & 0.07425 \end{bmatrix}.
\end{aligned}
$$

UNC/FORTH web archive

Online repository of models, tools, and traces

- Packet header, SNMP, SYSLOG, synthetic traces, …

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

Free login/ password to access it

☞ Simulation & emulation testbeds that replay synthetic traces for various traffic conditions

Mobile Computing Group @ University of Crete/FORTH
http://www.ics.forth.gr/mobile/

maria@csd.uoc.gr

Τέλος Ενότητας

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης