XIII_Workload Characterization for the web

Workload
Set of all inputs during any given period
Depends on the purpose of the study.

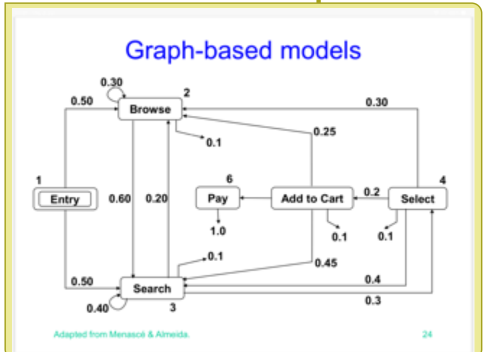
Basic component of a workload
refers to generic unit of work from the external.
Workload model is representation
that mimics the workload under study.

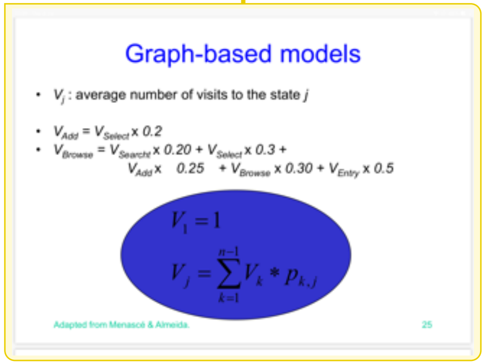
-**Description:**
Business -> user-oriented, such as #employees, invoices etc.
Functional -> describes programs, commands and requests.
Resource-oriented -> describes consumption of sys resources by the workload

See web server example

Refinement based on difference between documents size

-**Models:**
Natural: using basic components or traces of real workload
Artificial: don’t use anything of the real system (executable & non-executable)





Workload characterization methodology

1) Choice of an analysis standpoint
2) Identification of the basic component
3) Choice of the characterizing parameters
4) Data collection : assigns values to components
5) Partitioning the workload : real workloads can be viewed as a collection of heterogeneous components.
6) Calculating the class parameters

Workload partitioning:
-Resource usage
-Apps
-Objects
-Geographical orientation
-Functional
-Organizational units
-Mode

Calculate class parameters:
How?
- Averaging: average of parameters of all components.
- Clustering: a large number of components
are grouped into clusters of similar components.

New phenomena in the Internet and WWW

-**self-similarity:** bursty across several time scales.
-**heavy-tailed:** very large variability in the values
of the workload parameters.

Power Laws: $P[X>x]=kx^{-a}L(x)$

-**Burstiness modeling**
represented by (a,b)
-a is ratio between maximum observed request rate and avg of req
-b fraction of time arrival rate exceeds the average

Notation:
Consider L requests
tao: interval in consideration
Lambda: avg = L/tao
Epochs: n sub intervals = Tao/n
Arr(k) : #http requests that arrive in epoch k
Lambda_k: arrival rate in k

Arr^+ #req arrive in epochs in which $\text{lambda}_k > \text{lambda}$

b = (dimension of Arr+)/n

above-avg arrival rate, $\text{lambda}^+ = \text{Arr}^+/(\text{b*tao})$

$a = \text{lambda}^+/\text{lambda} = \text{Arr}^+/(\text{b*L})$

Impact of burstiness

Max throughput decreases as the burstiness factors increase.
To account for the burstiness effect, we write:
- $D = D_f + a*b$
- D_f portion independent from burstiness
- a used to inflate service demand according to $b = (U_1/X^1_0 - U_2/X^2_0)/(b1-b2)$

Prefer categorizing instead of averaging and multi class queuing network models with classes associated