Simulation: (study performance within parameter range)

easy to modify and update

possible to identify the optimal combination with simulation, but trade-off is not clear among different parameters

simulation is preferred beforehand due to the cost of changing system configuration

time consuming for a comprehensive simulation environment development

Measurement:(verify the analytical and simulation results if prototype exist)

only possible only if something similar to the proposed system already exists.

time: analytical modeling < measurements < simulations

What is system

hardware software

Evaluation techniques

measurement, simulation, analytical modeling

metrics to use

criteria used to quantify system performance

workload

the requests made by the users of the system

parameters(system/workload)

computer system:
a basic complete and functional computer, including all the hardware and software required to make it functional for a user. It should have the ability to receive user input, process data, and with the processed data, create information for storage and/or output.

analytical models are mathematical models that have a closed form solution(例子： the solution to the equations used to describe changes in a system can be expressed as a mathematical analytic function)

a set of mathematical equations such as:
Queueing models
Optimization models

Analytical modeling will be used to justify the consistency of measured values for different parameters.

simulation
easy to modify and update
Due to the cost of changing system configurations, simulation is preferred beforehand

with simulations, it may be possible to identify the optimal combination, but often it is not clear what the trade-off is among different parameters.
Time consuming for a comprehensive simulation environment development

Measurements are possible only if something similar to the proposed system already exists.
measurement is the least desirable technique in performance evaluation due to **highly variable environmental parameters.**
Saleability is the key justification when considering the expense and the labor of measurements.
Often used when you are given a system, such as software tester:
Unit testing
Integration testing
Smoke testing

analytical modeling and simulation can be used for situations where measurements is not possible, but in general it would be more convincing to others if the analytical modeling or simulation is based on previous measurement.

If results are required yesterday, the analytical modeling is probably the only choice.

The next consideration is the availability of tools: modeling skills, simulation languages, and measurement instruments.
Another consideration: level of accuracy desired

> analytical modeling requires many simplifications and assumptions that if the results turn out to be accurate.
> simulations can incorporate more details and require less assumptions than analytical modeling and, thus, more often are closer to reality.
> measurements, although they sounds like the real thing, may not give accurate results simply because many of the environmental parameter.

结合所有的评估方法
analytical model: find interesting range of parameters
simulation: study performance within parameter range
measurement: verify the analytical and simulation results if prototype exists

metrics:
Nominal capacity: maximum achievable under ideal conditions
Throughput: request/unit time
Usable capacity: max throughput for given response time limit
Efficiency: usable capacity/nominal capacity
Utilization: resource busy time/duration
Reliability: probability of errors
Availability: system shutdown time
Scalability: number of items the system can cope with

higher better: throughput
lower better: response time
nominal is best: Utilization(太高: low response time 太低:resource underused)

Quality of Service: = (throughput * correct transmission rate) / average request delay
global metrics
from the service provider's perspective

Quality of user Experience: = Average Round Trip Time * Average Allocated Bandwidth
individual users, most of the time
from the users' perspective

selecting metrics:

low variability: helps reduce the number of repetitions required to obtain a given level of statistical confidence.
Non-redundancy: If two metrics gives essentially the same information, it is less confusing to study only one.
Completeness:The set of metrics included in the study should be complete.

The workload consists of a list of service requests to the system.
it is essential that the workload be representative of the system usage in real life.
to produce representative workloads, one needs to measure and characterize the workload on existing system.

the workload for comparing several database systems may consist of a set of queries.

Real workload
- one observed during normal system operations
- Non-repeatable

Synthetic workload
- approximation of real workload
- can be applied repeatedly in a controlled manner
- no large data files, no sensitive data
- easily modified and ported
- easily measured

system parameters
- i**nclude both hardware and software parameters**, which generally do not vary among various installations of the system.

workload parameters
- may not be complete. That is after the first pass of the analysis, you may discover that there are additional parameters that affect the performance.

经常变换的参数 factor 他们的数值叫value

最好从少部分的factor和较小的level开始

for each of these two factors you may choose only two levels: small and large

some of the factors can be set at fixed values at the initial experimental stage

system under study may not be available

simulation may be preferred alternative to environment
- controlled study of wider range of workloads and environments

higher accuracy results than analytical modeling

Sometime simulation is not preferred
accurate simulation models takes a long time to develop

simulation容易修改，更新
因为系统结构也尝尝变化，simulation is preferred beforehand

用simulation可能去找出最好的组合，但是很难在各种参数之间找到平衡(trade-off)
容易理解的模拟环境非常耗时

evaluation的好技巧：
Combining evaluation techniques is useful
analytical model: find interesting range of parameters
simulation: study performance within parameter range
到验证之前，所有的评估结果都不可信
always validate one analysis modality with another
beware of counterintuitive results

Common mistakes: too much detail
level of detail limited only by time available for development
a detailed model may not be a better model并不是越细致越好

recipe for success
1.start with less-detailed model
2. get some results
3.study sensitivities
4.introduce details in key areas that affect results most

Common mistakes: initial conditions
Initial part of a simulation 不具有代表性，需要被discarded
transient behavior rather than steady state
several techniques for identifying beginning of steady state

Common mistakes: too short simulations
Simulation run times 通常很长
Temptation is to halt simulations ASAP
但是通常结果都会严重依赖于初始条件
可能不会很有代表性，直到到达steady state
correct length for simulations depends on accuracy desired(width of confidence intervals)
Variance of observed quantities

Common mistakes: bad random numbers
不好的随机数可能会污染simulation results
1.period too short
2.assume global randomness = local randomness
3.rely on bit subsets: may not be as random as whole
Use well-known generator rather than rolling your own

simulation type:

Monte Carlo Simulation
Model probabilistic phenomenon that do not change over time
It is used for evaluating non probabilistic expressions using probabilistic methods.(赌博那个例子)

Trace-Driven Simulation
the model's inputs are derived from a sequence of observations made on a real system.
Trace = time ordered record of events on real system
Statistics(requests, errors, latency, etc.) are calculated based on the full volume of traces.
eg:total requests and requests per second, latency, breakdown of time spent by service/type

优势:

credibility, easy validation, accurate workload, detailed trade-offs, less randomness, fair comparison, similarity to the actual implementation
缺点:
Complexity, representativeness, finiteness, single point of validation, detail, trade-off

Discrete-event simulations
use discrete-state model of system, the simulation will have the following components

mostly used in CS
time driven system: auto-sync programs, alert systems, some street or home lighting systems
event driven system: almost every computer systems

What are the three common performance evaluation techniques and when to use them? Talk a little bit about their advantages and disadvantages.

measurement: only possible if something similar to the proposed system already exists.
It may not give accurate results simply because many of the highly variable environmental parameter

analytical modeling: sometimes results are required yesterday, then analytical modeling is the only choice. It will be used to justify the consistency of measured values for different parameters.

simulation: easy to update and modify
simulation is preferred beforehand, due to the cost of changing system configuration
simulation may be preferred alternative to measurement(controlled study of wider range of workloads and environments)
higher accuracy results than analytical modeling
With simulation, it is possible to identify the optimal combination, but often it is not clear what the trade-off is among different parameters. Accurate simulation models takes a long time to develop. Simulations can incorporate more details and require less assumptions than analytical modeling and, thus, more often are closer to reality.

What will be the main concerns when deciding which simulator to use for system evaluation? List four of those you think are important.
accuracy, cost, easy to use, time required, trade-off evaluation, saleability, tools, stage

Random number generation is often required to generate synthetic data. Linear-Congruential Generators are the popular ones that can be applied efficiently. List at least three factors that you can use to determine whether a random number sequence is good or not.
the period of random number sequence must be long

the time of generating random number sequence must be short

The correlation between successive numbers should be small

Explain what is workload in your own words. In the LEACH evaluation, what has been used as workload to test the performance.
The requests made by the users of the system
The workload consists of a list of service requests to the system

In simulation, random number generation is open required. Linear-Congruential Generators are the popular ones that can be applied efficiently. Explain how can you obtain a full period generator.

$x_n = (ax_{n-1} + b) \bmod m$
modulus m should be large
m = 2k, for some integer k
a=4c+1, for some integer c
b is an odd integer

What are the differences between factors and parameters? Present the answer in your own words and also give some examples for both.

Some parameters can vary and some are fixed.
The parameters to be varied are called factors and their values are called levels.

Workload Selection is essential for performance evaluation. Give your ideas and opinions on workload selection in terms of level of details.

Most frequent request
- valid if one service is requested much more often than others
- examples: add instruction, kernels

Frequency of request types
- example: instruction mix
- context sensitive services - must use a set (e.g. caching)

Time stamped sequence of requests (trace)
- too much detail for analytical modeling
- may require exact reproduction of component behavior for timing

Average resource demand
- analytical models use request rate rather than requests
- group similar services in classes; use avg. demand per class

Distribution of resource demands, used if
- variance in resource demands is large
- distribution impacts performance

A simple graphical method for comparing two sets of sample quantiles.
Sample quantiles are based on order statistics
Q–Q plots are commonly used to compare a data set to a theoretical model.
- goodness of fit
- Difference between measured and predicted values is modeling error

Q–Q plots are also used to compare two theoretical distributions to each other.

a. A.Monte Cario simulation is specially suitable for one single metric evaluation scenario.(对)

b. B.Long run is a good strategy to use to avoid unstable state in simulation(不是，因为过于浪费

资源了吧)

c. C.Normally when using Measurement as a method for system evaluation, we do not need a second one since it already has acceptable accuracy and high salability(过于绝对)

d. D.In many cases, invalid but verified simulation is better than valid but unverified simulation(验证过的可靠一些)