Steps in Performance Evaluation

Tran, Van Hoai (hoai@hcmut.edu.vn)

Faculty of Computer Science & Engineering
HCMC University of Technology

2022-2023/Semester 1

Outline

1 Systematic approach to performance evaluation (PE)

2 Common mistakes

3 Home exercise

Outline

1 Systematic approach to performance evaluation (PE)

2 Common mistakes

3 Home exercise

The unique characteristics

- **Metrics**: the criteria used to evaluate the performance of the system. components.
- **Workloads**: the requests made by the users of the system.

- Power consumption
- Number of queries sent to a database
- Compression ration of a data storage
- Throughput (rate of production)
- The interval between HTTP requests to a web server
- CPU utilization
- Volume of clean water used per household
- SPECfp
- LinPack

The unique characteristics

- **Metrics**: the criteria used to evaluate the performance of the system. components.
- **Workloads**: the requests made by the users of the system.

- Power consumption (metric)
- Number of queries sent to a database (a workload property)
- Compression ration of a data storage (metric)
- Throughput (rate of production) (metric)
- The interval between HTTP requests to a web server (a workload property)
- CPU utilization (metric)
- Volume of clean water used per household (metric)
- SPECfp (workload)
- LinPack (workload)

The unique characteristics

- **Techniques**: a methodology to collect the metrics under the given workloads.

The unique characteristics

- **Techniques**: a methodology to collect the metrics under the given workloads.

Evaluating the performance of Internet to website

The unique characteristics

- **Techniques**: a methodology to collect the metrics under the given workloads.

Evaluating the performance of Internet to website

- Which techniques should be used?

The unique characteristics

- **Techniques**: a methodology to collect the metrics under the given workloads.

Evaluating the performance of Internet to website

- Which techniques should be used?
- Which metrics to be evaluated?

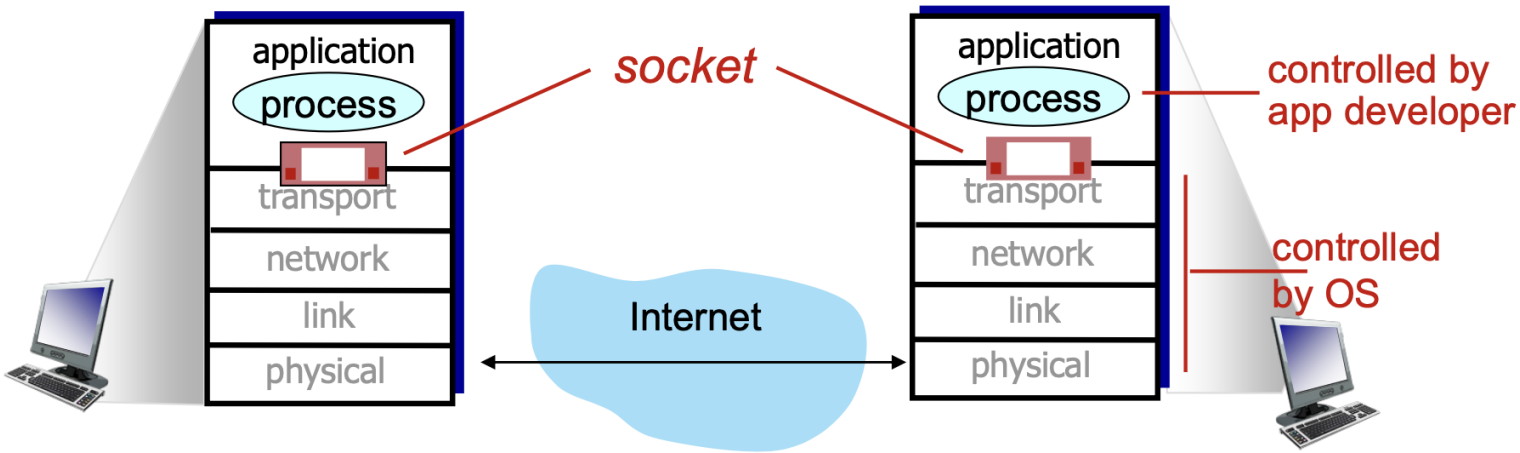
The unique characteristics

- **Techniques**: a methodology to collect the metrics under the given workloads.

Evaluating the performance of Internet to website

- Which techniques should be used?
- Which metrics to be evaluated?
- Which workloads to be applied to the PE system ?

Performance evaluation of a website



- Which techniques should be used? (`ping`, `curl`)
- Which metrics to be evaluated? (round-trip communication, bandwidth, error,...)
- Which workloads to be applied to the PE system ? (bursty, interactive, latency sensitive, non-realtime)

Steps in PE project

1. State goals and Define the System
2. List Services and Outcomes
3. Select Metrics
4. List Parameters
5. List Factors to Study
6. Select Evaluation Technique
7. Select Workload
8. Design Experiments
9. Analyze and Interpret Data
10. Present Results

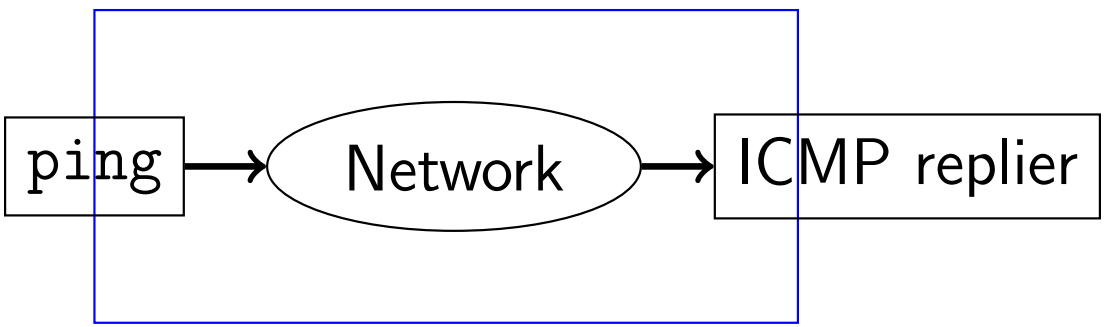
Step 1: State Goals and Define the System

- Delineate (describe) the system boundaries.
- System definition may depend on the study goals.

Step 1: State Goals and Define the System

- Delineate (describe) the system boundaries.
- System definition may depend on the study goals.

- Goal 1 = "to estimate their impact on the response time of interactive users".
  $\Rightarrow$ System = "time sharing + CPUs + other computer components"
- Goal 2 = "to decide which ALU"
  $\Rightarrow$ System = "Components of CPUs"



Step 1: System boundary example

System boundary

The system boundary is a conceptual line that divides the system that you want to study from 'everything else'.

Which of the following is correct system boundary of ATM ?

- The physical machine itself
- The customer and the machine
- The machine and the bank's local database of customer accounts
- The machine and the bank's national network and central database of transactions and account balances
- The machine and the staff that loan and run regular software checks of the machine

Different analyses may have different boundaries.

Step 2: List Services and Outcomes

- A system always provides a set of services.
- Each service has its own outcomes.
  All possible outcomes of a service must be listed.

Step 2: List Services and Outcomes

- A system always provides a set of services.
- Each service has its own outcomes.
  All possible outcomes of a service must be listed.

- A service of computer network = "to send packets to specified destination"
- A service of CPU scheduler = "to schedule jobs for CPU execution"
- A service of database = "to response to queries"
  $\Rightarrow$ Possible outcomes: answer correctly; incorrectly; not at all; ...

Step 2: List Services and Outcomes

- A system always provides a set of services.
- Each service has its own outcomes.
  All possible outcomes of a service must be listed.

- A service of computer network = "to send packets to specified destination"
- A service of CPU scheduler = "to schedule jobs for CPU execution"
- A service of database = "to response to queries"
  $\Rightarrow$ Possible outcomes: answer correctly; incorrectly; not at all; ...

List of services and possible outcomes are useful in selecting right metrics and workloads.

Step 3: Select Metrics

- Depending on the listed outcomes
- Metrics are related to the speed, accuracy, and availability of services.

Step 3: Select Metrics

- Depending on the listed outcomes
- Metrics are related to the speed, accuracy, and availability of services.

- (Computer network) An outcome "No errors and failures" $\Rightarrow$ round-trip time; bandwidth;..."
- (Computer network) An outcome "Some errors" $\Rightarrow$ error rate;..."
- Metrics of processor = "speed (time taken to execute) of various instructions;..."

Step 3: Select Metrics

- Depending on the listed outcomes
- Metrics are related to the speed, accuracy, and availability of services.

- (Computer network) An outcome "No errors and failures" $\Rightarrow$ round-trip time; bandwidth;..."
- (Computer network) An outcome "Some errors" $\Rightarrow$ error rate;..."
- Metrics of processor = "speed (time taken to execute) of various instructions;..."

How to choose correct metrics?

- In this lecture, just list all possible metrics (as you recognize).
- Choosing correct metrics: Lecture #3, maybe :-)

Step 4: List Parameters

- List all possible parameters that affect performance.
- Parameter types:
  - **System parameter**: not vary much among various system installations.
  - **Workload parameter**: vary from one installation to the next.
- Parameter set should be comprehensive (parameters added incrementally)

Step 4: List Parameters

- List all possible parameters that affect performance.
- Parameter types:
  - **System parameter**: not vary much among various system installations.
  - **Workload parameter**: vary from one installation to the next.
- Parameter set should be comprehensive (parameters added incrementally)

System parameters

- Speed of local and remote CPUs
- Network speed
- Operating systems involved in the system

Workload parameters

- Time between successive `ping` send
- Number of bytes in a `ping` packet
- Time-to-live (ttl) for a packet
- Other loads on the system

Step 4: List Parameters

- List all possible parameters that affect performance.
- Parameter types:
  - **System parameter**: not vary much among various system installations.
  - **Workload parameter**: vary from one installation to the next.
- Parameter set should be comprehensive (parameters added incrementally)

System parameters

- Speed of local and remote CPUs
- Network speed
- Operating systems involved in the system

Workload parameters

- Time between successive `ping` send
- Number of bytes in a `ping` packet
- Time-to-live (ttl) for a packet
- Other loads on the system

- Helpful to determine what data needs to be collected before or during the analysis.

Step 5: List Factors to Study

- Parameters to be varied are called **factors**.
- Different values of factors are called **levels**.
- Parameters which has high impact on performance are preferably selected as factors.
- Gradually extend factor list and per-factor levels

Step 5: List Factors to Study

- Parameters to be varied are called **factors**.
- Different values of factors are called **levels**.
- Parameters which has high impact on performance are preferably selected as factors.
- Gradually extend factor list and per-factor levels

- Transmission media: copper cable, fiber cable, wireless,
- Size of network: short or long distance
- Packet size: set of levels $= \{0, 64, 128, ..., 4096\}$

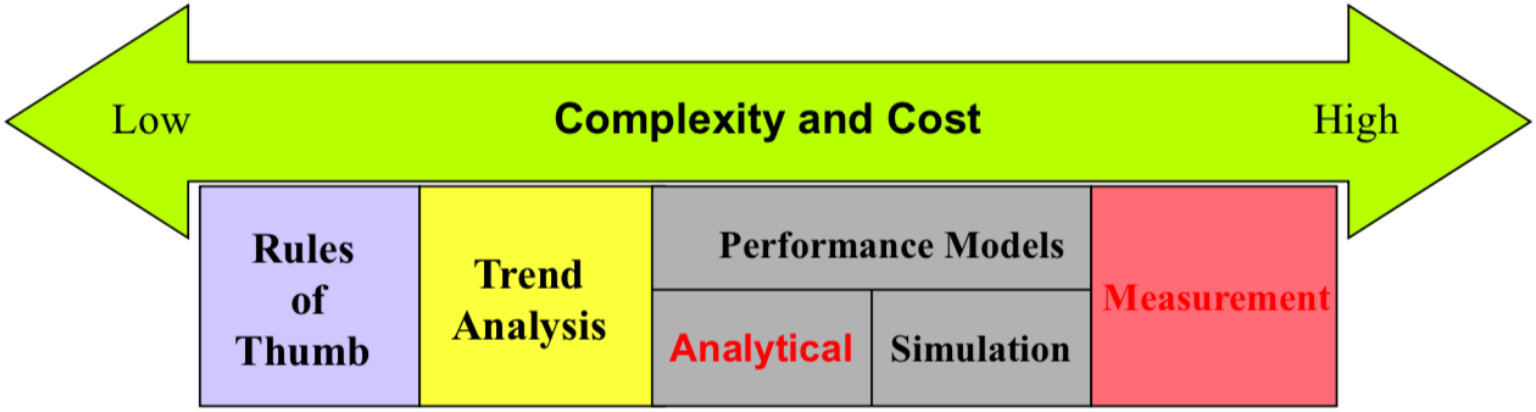
Step 5: List Factors to Study

- Parameters to be varied are called **factors**.
- Different values of factors are called **levels**.
- Parameters which has high impact on performance are preferably selected as factors.
- Gradually extend factor list and per-factor levels

- Transmission media: copper cable, fiber cable, wireless,
- Size of network: short or long distance
- Packet size: set of levels $= \{0, 64, 128, ..., 4096\}$

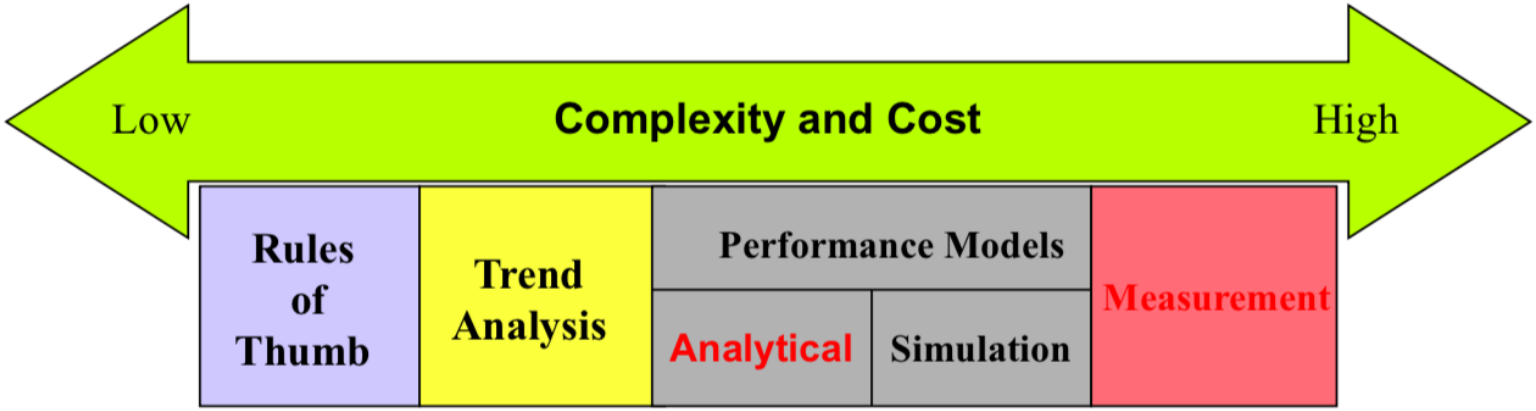
- Selected factors must be feasible (*khả thi*) to decision makers' control.
- Non-factor parameters must be fixed (and also feasible and low impact).

Step 6: Select Evaluation Technique



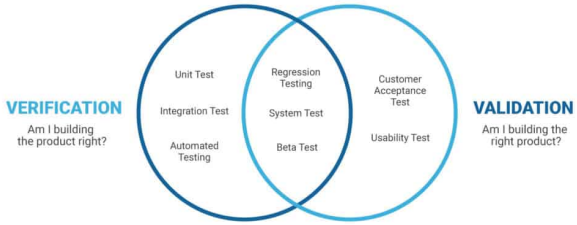
- Selection of right techniques also depends on desired level of accuracy.

Step 6: Select Evaluation Technique

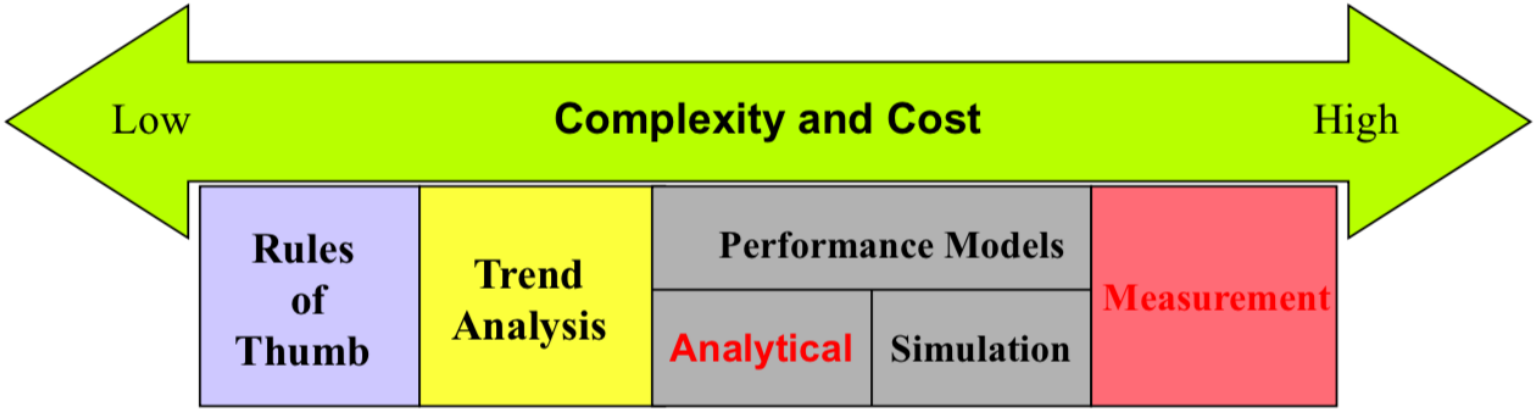


- Selection of right techniques also depends on desired level of accuracy.

- Prototypes implemented ⇒ measurement.
- Use analytical modeling for validation.

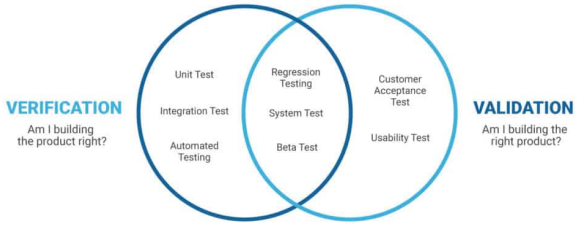


Step 6: Select Evaluation Technique



- Selection of right techniques also depends on desired level of accuracy.

- Prototypes implemented ⇒ measurement.
- Use analytical modeling for validation.



- Choosing suitable techniques: Lecture #3, maybe :-)

Step 7: Select Workload

- Workload = list of service requests
- Workload $\approx$ system usage in real life
- Expressing workload depends upon evaluation technique.

Step 7: Select Workload

- Workload = list of service requests
- Workload $\approx$ system usage in real life
- Expressing workload depends upon evaluation technique.

- Analytical modeling: a probability of various requests.
- Simulation: a trace of requests measured on a real system.
- Measurement: user scripts to be executed on the systems.

Step 7: Select Workload

- Workload = list of service requests
- Workload $\approx$ system usage in real life
- Expressing workload depends upon evaluation technique.

- Analytical modeling: a probability of various requests.
- Simulation: a trace of requests measured on a real system.
- Measurement: user scripts to be executed on the systems.

- Choosing suitable workloads: Lecture #4, maybe :-)

Step 8: Design Experiments

- Based on list of factors and their levels, deciding a sequence of experiments that offer
  - maximum information: co-relationship between factors and metrics,
  - minimal effort: number of experiments.
- 2-phase approach
  1. Phase 1: large number of factors, small number of levels.
  2. Phase 2: reduced number of factors, increased number of levels.

Step 8: Design Experiments

- Based on list of factors and their levels, deciding a sequence of experiments that offer
  - maximum information: co-relationship between factors and metrics,
  - minimal effort: number of experiments.
- 2-phase approach
  1. Phase 1: large number of factors, small number of levels.
  2. Phase 2: reduced number of factors, increased number of levels.

Factors below are considered to compare remote pipe and remote procedure call (RPC).

- Type of channels: Remote pipes and remote procedure calls (2 levels).
- Size of the Network: Short distance and long distance (2 levels).
- Sizes of the call parameters: small and large (2 levels).
- Number n of consecutive calls=Block size: 1, 2,. . . , 1024 (11 levels).

$\Rightarrow$ # experiments (full factorial) $= 2^3 \times 11 = 88$.

Step 9: Analyze and Interpret Data

- Outcomes of measurements and simulations are random quantities.
  $\Rightarrow$ Both mean and variability are investigated.
- Analysts work on results, decision makers draw conclusions.

Step 9: Analyze and Interpret Data

- Outcomes of measurements and simulations are random quantities.
  $\Rightarrow$ Both mean and variability are investigated.
- Analysts work on results, decision makers draw conclusions.

Using the same example in Step 8.

- Analysis of Variance (ANOVA) for the first three factors.
- Regression for number n of successive calls.

Step 10: Present Results

- Presentation of results should be understandable by decision makers.
  $\Rightarrow$ Graphic graph plotting.
- It is possible to go back to previous steps.

Step 10: Present Results

- Presentation of results should be understandable by decision makers.
  $\Rightarrow$ Graphic graph plotting.
- It is possible to go back to previous steps.

- Trend of results plotted as a function of n.
- Results are strongly fluctuated $\Rightarrow$ system boundary redefinition.
- Remote pipe = RPC $\Rightarrow$ workload is too light?.

Step 10: Present Results

- Presentation of results should be understandable by decision makers.
  $\Rightarrow$ Graphic graph plotting.
- It is possible to go back to previous steps.
  - Trend of results plotted as a function of n.
  - Results are strongly fluctuated $\Rightarrow$ system boundary redefinition.
  - Remote pipe = RPC $\Rightarrow$ workload is too light?.
- Reading Chapter 10 in textbook, maybe or low score :-)

Outline

1 Systematic approach to performance evaluation (PE)

2 Common mistakes

3 Home exercise

Basic subset of common mistakes

Students are advised to experience themselves.

1. No goals: goals $\Rightarrow$ metrics, workloads, techniques
2. Biased goals: "to show that OUR system is better than THEIRS"
3. Unsystematic approach
4. Analysis Without Understanding the Problem
5. Incorrect Performance Metrics
6. Unrepresentative Workload
7. Wrong Evaluation Technique
10. Inappropriate Experimental Design
12. No Analysis
18. Ignoring Variability
20. Improper Presentation of Results

Outline

1 Systematic approach to performance evaluation (PE)

2 Common mistakes

3 Home exercise

A report on PE project

To compare (w.r.t. performance) different methods to find shortest paths in Python

- 3 methods: Dijkstra, Bellman-Ford, Johnson.
- Briefly describe (along with justification (*lý giải*)) the system and list:
  - Services
  - Performance metrics
  - System parameters
  - Workload parameters
  - Factors and their ranges
  - Evaluation technique
  - Workload

If a student submits a report (4 page long) within 2 weeks, his/her lab score will be bonused 5 points (/10).