Autonomous Networking

Gaia Maselli
Dept. of Computer Science

Today's plan

How do we evaluate the performance of a new protocol?

- Selection of an evaluation technique
- Selection of performance metrics
- Results representation

A systematic approach to performance evaluation

1. **State Goals and Define the System**
   - Compare a new MAC protocol for sensor networks with an existing one
2. **List Services and Outcomes**
   - Send packets to a specific node called sink
   - Success or failure in packet delivery
3. **Select Metrics**
   - Packet delivery ratio, energy consumption, network lifetime
4. **List Parameters**
   - number of nodes, duty cycle, transmission energy cost, etc.
5. **Select Factors to Study**
   - Message inter-arrival period, duty cycle

A systematic approach to performance evaluation

SAPIENZA UNIVERSITÀ DI ROMA

6. **Select Evaluation Technique** (depends upon the time and resources available)
   - simulation
7. **Select Workload**
   - number of data flows in the network, packets generated per unit of time
8. **Design Experiments**
   - Decide a sequence of experiments that maximize information with minimal effort
9. **Analyze and Interpret Data**
   - Draw conclusions from results
10. **Present Results**
    - Make graphs (ex., plot energy consumption by varying message inter arrival time)

Repeat (go back and reconsider some decisions)

Selecting an evaluation technique

Selecting an evaluation technique

- Three techniques for performance evaluation
  1. Analytical modeling
  2. Simulation
  3. Measurement

- How do we choose one of them?

Criteria for Selecting an Evaluation Technique

Criterion		Analytical Modeling	Simulation	Measurement
1.	Stage	Any	Any	Postprototype
2.	Time required	Small	Medium	Varies
3.	Tools	Analysts	Computer languages	Instrumentation
4.	Accuracy[a]	Low	Moderate	Varies
5.	Trade-off evaluation	Easy	Moderate	Difficult
6.	Cost	Small	Medium	High
7.	Saleability	Low	Medium	High

[a] In all cases, result may be misleading or wrong.

Criterion 1: Stage

- The key consideration in deciding the evaluation technique is the *life-cycle stage* in which the system is

- New system ➔ *analytical modeling* and *simulation* are the only techniques from which to choose

- Prototype or Improved system

  ➔ *measurement* (but also modeling and simulation)

Modeling and simulation can be done anytime, measurement requires a prototype

Criteria for Selecting an Evaluation Technique

Criterion		Analytical Modeling	Simulation	Measurement
1.	Stage	Any	Any	Postprototype
2.	Time required	Small	Medium	Varies
3.	Tools	Analysts	Computer languages	Instrumentation
4.	Accuracy[a]	Low	Moderate	Varies
5.	Trade-off evaluation	Easy	Moderate	Difficult
6.	Cost	Small	Medium	High
7.	Saleability	Low	Medium	High

[a] In all cases, result may be misleading or wrong.

Criterion 2: Time required

- Time available for evaluation has to be taken into account
- Often results are required *yesterday*

Short	➔	*analytical modeling*
Medium	➔	*simulation*
Long	➔	*measurement*

Murphy's law strikes measurements more often than other techniques

↓

Variable time

Criteria for Selecting an Evaluation Technique

Criterion		Analytical Modeling	Simulation	Measurement
1.	Stage	Any	Any	Postprototype
2.	Time required	Small	Medium	Varies
3.	Tools	Analysts	Computer languages	Instrumentation
4.	Accuracy[a]	Low	Moderate	Varies
5.	Trade-off evaluation	Easy	Moderate	Difficult
6.	Cost	Small	Medium	High
7.	Saleability	Low	Medium	High

[a] In all cases, result may be misleading or wrong.

Criterion 3: Tools

- Availability of tools plays an important role

Modeling skills

Simulation languages

Measurement instruments

Criteria for Selecting an Evaluation Technique

Criterion		Analytical Modeling	Simulation	Measurement
1.	Stage	Any	Any	Postprototype
2.	Time required	Small	Medium	Varies
3.	Tools	Analysts	Computer languages	Instrumentation
4.	Accuracy[a]	Low	Moderate	Varies
5.	Trade-off evaluation	Easy	Moderate	Difficult
6.	Cost	Small	Medium	High
7.	Saleability	Low	Medium	High

[a] In all cases, result may be misleading or wrong.

Criterion 4: Accuracy

- Level of accuracy desired is another important consideration

- **Low ➔** ***analytical modeling*** may require so many simplifications and assumptions that results may be too approximate
- **Moderate ➔** ***simulations*** *can* incorporate more details and and require less assumptions and thus are closer to reality
- **Variable ➔** ***measurement*** may not give accurate results simply because many of the environmental parameters, such as system configuration, type of workload, and time of measurement may be unique to the experiment

Criteria for Selecting an Evaluation Technique

Criterion		Analytical Modeling	Simulation	Measurement
1.	Stage	Any	Any	Postprototype
2.	Time required	Small	Medium	Varies
3.	Tools	Analysts	Computer languages	Instrumentation
4.	Accuracy[a]	Low	Moderate	Varies
5.	Trade-off evaluation	Easy	Moderate	Difficult
6.	Cost	Small	Medium	High
7.	Saleability	Low	Medium	High

[a] In all cases, result may be misleading or wrong.

Criterion 5: Trade-off evaluation

- The goal of every performance study is to compare different alternatives or to find the optimal parameter values

Easy	→	*analytical modeling*
Moderate	→	*simulation*
Difficult	→	*measurement*

Criteria for Selecting an Evaluation Technique

Criterion		Analytical Modeling	Simulation	Measurement
1.	Stage	Any	Any	Postprototype
2.	Time required	Small	Medium	Varies
3.	Tools	Analysts	Computer languages	Instrumentation
4.	Accuracy[a]	Low	Moderate	Varies
5.	Trade-off evaluation	Easy	Moderate	Difficult
6.	Cost	Small	Medium	High
7.	Saleability	Low	Medium	High

[a] In all cases, result may be misleading or wrong.

Criterion 6: Cost

- Cost allocated for the project is also important

- Small ➔ *analytical modeling* requires only paper and pencil (in addition to the analyst's time)

- Medium ➔ *simulation* requires a simulator (often free) and some time

- High ➔ *measurement* requires real equipment, instruments and time

Criteria for Selecting an Evaluation Technique

Criterion		Analytical Modeling	Simulation	Measurement
1.	Stage	Any	Any	Postprototype
2.	Time required	Small	Medium	Varies
3.	Tools	Analysts	Computer languages	Instrumentation
4.	Accuracy[a]	Low	Moderate	Varies
5.	Trade-off evaluation	Easy	Moderate	Difficult
6.	Cost	Small	Medium	High
7.	Saleability	Low	Medium	High

[a] In all cases, result may be misleading or wrong.

Criterion 7: Saleability

- Saleability is the key justification when considering the expense and labor of measurements.

- Low → *analytical modeling* - some people are skeptical of analytical results simply because they do not understand the technique or the final results

- Medium → *simulation*

- High → *measurement* - It is easy to convince others if it is a real measurement

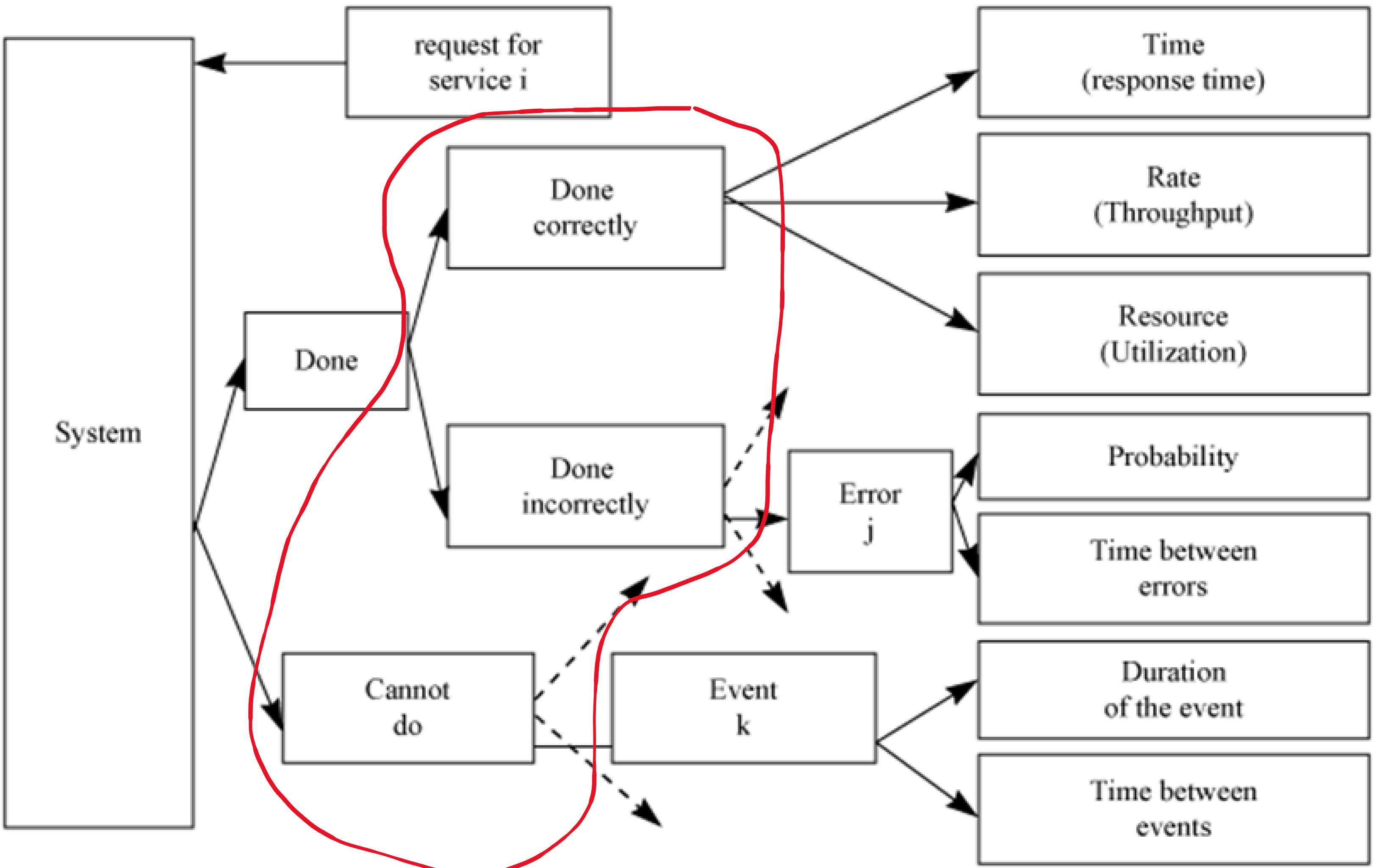
Which technique do we choose?

Three rules of validation

- Do not trust the results of a **simulation model** until they have been validated by analytical modeling or measurements
- Do not trust the results of an **analytical model** until they have been validated by simulation model or measurements
- Do not trust the results of a **measurement** until they have been validated by analytical modeling or simulation

Selecting performance metrics

Selecting performance metrics

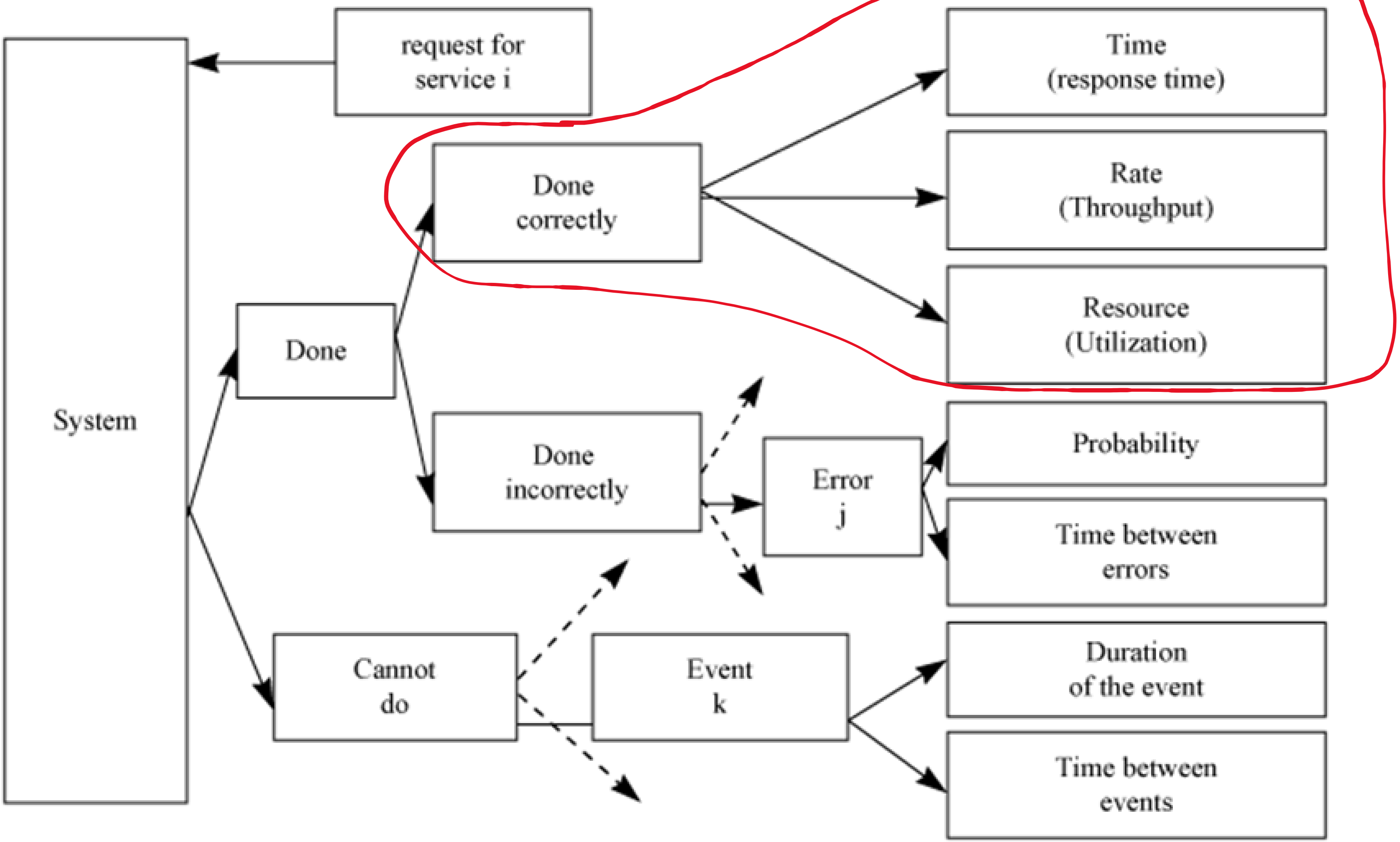


Selecting performance metrics

- For each service request the system may **perform the service**
  - **Correctly**
  - **Incorrectly**
  - **Refuse** to perform the service

Example: a gateway in a network offers the service of forwarding packets to the specified destination. When presented a packet, it may forward the packet correctly, it may forward it to the wrong destination, or it may be down (not forward it at all)

Selecting performance metrics



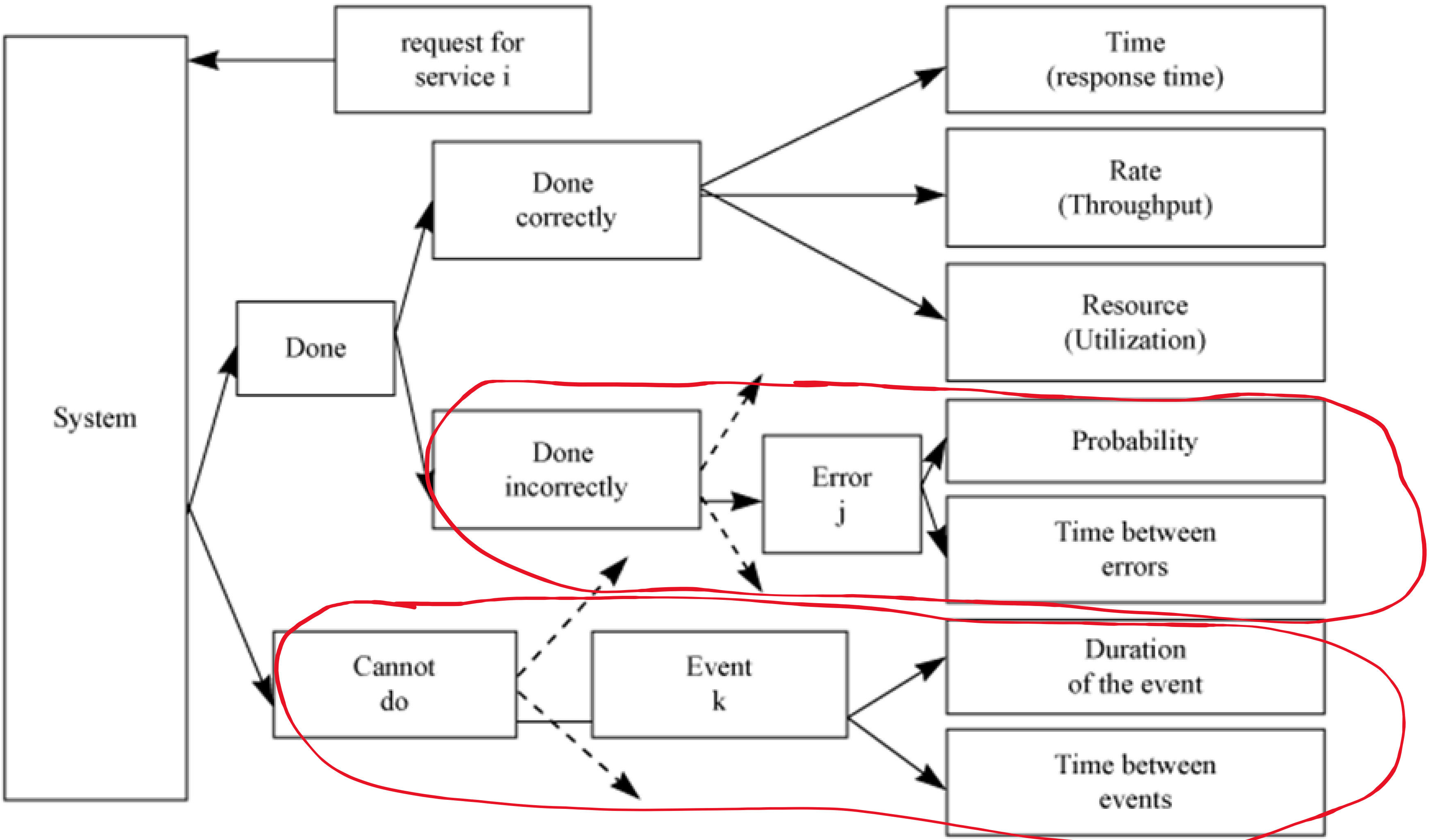
Selecting performance metrics

- If the systems performs the service correctly, its performance is measured by
  - the **time taken** to perform the service (responsiveness)
  - **the rate at which** the service is performed (productivity)
  - **the resource consumed** while performing the service (utilization)

Example (gateway):

- Responsiveness is the time interval between arrival of a packet and its successful delivery
- Productivity is the number of packets forwarded per unit of time
- Utilization is percentage of time gateway resources are busy for the given load level

Selecting performance metrics



Selecting performance metrics

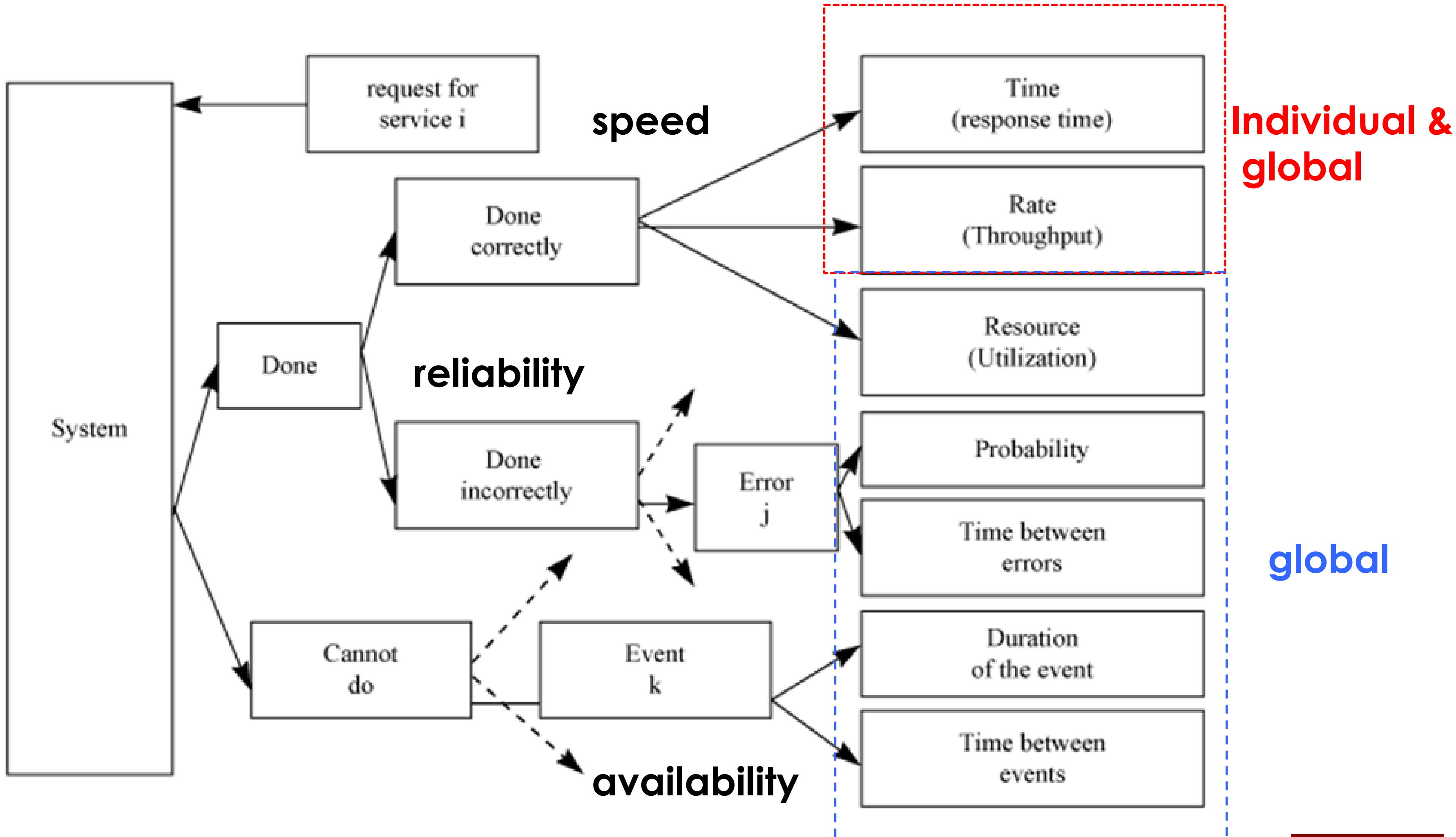
1. If the system performs the service incorrectly, an error is occurred
   - It is helpful to classify errors and to determine the probabilities of each class of errors.

     Example (gateway): we may want to find the probability of single-bit errors and packet error

2. If the system does not perform the service, it is said to be down, failed or unavailable
   - It is helpful to classify the failure modes and to determine the probability of each class

     Example (gateway): the gateway may be unavailable 0.01% of the time due to processor failure and 0.03% due to software failure

Selecting performance metrics



Selecting performance metrics

- Names of the metrics associated with the three outcomes
  - **successful service ➔ speed**
  - **Error ➔ reliability**
  - **Unavailability ➔ availability**
- For each service offered by the system, one would have a number of speed metrics, a number of reliability metrics, and a number of availability metrics
- Most systems offer more than one service, and thus the number of metrics grows proportionally
- As a network is shared by multiple users, two types of performance metrics need to be considered**: individual** and **global**
  - **Individua**l metrics reflect the **utility of each user**
  - **Globa**l metrics reflect the **system wide utility**
  - Some metrics are **individual and global**

N.B. there are cases when the decision that optimize individual metrics is different from the one that optimizes the system metric (e.g., throughput !!!)

Selecting performance metrics

- Given a number of metrics, use the following considerations to select a subset:
- **Non redundancy**
  - If two metrics give essentially the same information, it is less confusing to study only one
- **Completeness**
  - All possible outcomes should be reflected in the set of performance metrics

Case study

1. We design a new MAC protocol for WSN
2. We want to evaluate our protocol
3. We select an evaluation techniques
    1. Simulation
4. We implement our protocol in the simulator
5. We want to show that it is a great protocol (compare with state of the art protocols)
6. We set a scenario (number of nodes) and workload

Case study

7. We decide metrics
   1. Time (packet delay)
   2. Rate (Throughput)
   3. Resources (energy consumed)
   4. …

8. We run simulations multiple times (repetitions, making some parameters to vary)

9. We got many values… how do we present results?

Summarizing measured data

Summarizing measured data

SUMMARIZING DATA BY A **SINGLE** NUMBER

SUMMARIZING **VARIABILITY**

Motivation

- A measurement project may result in several hundreds or millions of observations on a given variable.
- To present the measurements it is necessary to summarize data
- **How to report the performance as a single number**?
- **Is specifying the mean the correct way**?
- How to report the variability of measured quantities? What are the alternatives to variance and when are they appropriate?

Summarizing data by a single number: central tendency

Central tendency: empirical mean

- The simplest description of numerical data
- Given a dataset $\{x_i, i=1,\ldots,N\}$ central tendency tells us where on the number line the values tend to be located
- ***Empirical mean (average) or sample mean*** is the most widely used measure of central tendency
- It is obtained by taking the sum of all observations and dividing this sum by the number of observations in the sample.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Common misuses of means

Using mean of significantly different values

- There are cases when the mean is not useful
- ***Usefulness*** depends upon **the number of values and the variance**
- Example: 2 measurements for packet latency
  - l1 = 2ms
  - l2 = 5 s
  - Mean packet latency = 2,501s
  - An analysis based on 2,501s would lead nowhere close to the two possibilities (the mean is the correct index but is useless)

Common misuses of means (Cont)

Using mean without regard to skewness of distribution

System A	System B
10	5
9	5
11	5
10	4
10	31
Sum=50	Sum=50
Mean=10	Mean=10
Typical=10	Typical=5

- Both systems have mean response times of 10
- System A: it is **useful** to know the mean because the variance is low and 10 is the typical value
- System B: the typical value is 5; hence using 10 for the mean does not give any useful result. The variability is too large in this case

Common misuses of means (Cont)

Multiplying means to get the mean of a product

- If the variable are correlated (not independent)

$$E(xy) \neq E(x)E(y)$$

Taking a mean of a ratio with different bases

Geometric mean

- The geometric mean of n values $x_1, x_2, \ldots, x_n$ is obtained by multiplying the values together and taking the nth root of the product

Example: The performance improvements in 7 layers.

What is the average improvement per layer?

Protocol Layer	Performance Improvement
7	18%
6	13%
5	11%
4	8%
3	10%
2	28%
1	5%

Geometric mean (Cont)

Example: The performance improvements in 7 layers.

What is the average improvement per layer?

Protocol Layer	Performance Improvement
7	18%
6	13%
5	11%
4	8%
3	10%
2	28%
1	5%

$$\text{Average improvement per layer}$$
$$= \{(1.18)(1.13)(1.11)(1.08)(1.10)(1.28)(1.05)\}^{\frac{1}{7}} - 1$$
$$= 0.13$$

Summarizing variability: dataset dispersion

Summarizing variability

- Given a data set, summarizing it by a single number is rarely enough
- It is important to include a statement about its **variability** in any summary of the data
- Motivation: given two systems with the same mean performance, one would generally prefer one whose performance does not vary much from the mean
- Systems with low variability are preferred
- Variability is specified using one of the following measures which are called *indices of dispersion*
  - Range minimum and maximum of the values observed
  - Variance and standard deviation
  - Percentiles and quantiles
  - Mean absolute variation

Range

- The range of a stream of values can be easily calculated by keeping track of the minimum and the maximum
- **Range = Max-Min**
- Larger range => higher variability
- In most cases, range is not very useful.
- The minimum often comes out to be zero and the maximum comes out to be an ``outlier'' far from typical values.
- Unless the variable is bounded, the maximum goes on increasing with the number of observations, the minimum goes on decreasing with the number of observations, and there is no “stable'' point that gives a good indication of the actual range.
- Range is useful if, and only if, there is a reason to believe that the **variable is bounded**.

Variance and coefficient of variation

- ***Variance*** is measured in squared units

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$\text{where } \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

- ***Standard deviation*** *is* the square root of variance and is expressed in the same units as the data

- **Coefficient of variation** is (standard deviation/mean) and is a dimensionless measure of the dispersion of a dataset

Quantile and percentile

- A more detailed description of dataset dispersion is in terms of quantiles and percentile
- The *pth quantile* is the value below which the *fraction p* of the values lie
- The median is the 0.5 quantile
- This can also be expressed as a *percentile*, e.g., the 90 th percentile is the value that is larger than 90% of the data
- Specifying the 5-percentile and the 95-percentile of a variable has the same impact as specifying its minimum and maximum

Quantile calculation

- The **a-quantile** can be estimated by sorting the observations and taking the **[(n-1)a + 1]-th element** in the ordered set.

Example: Given a set of 32 elements, we first order it, then

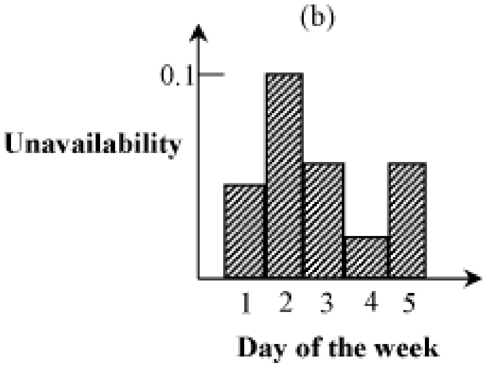
- 0.1-quantile or 10-percentile = (31*0.1)+1 = 4.1 = 4^{th} element
- 0.9-quantile or 90-percentile = (31*0.9)+1 = 28.9 = 29^{th} element
- First quartile = (31*0.25)+1 = 8.75 = 9^{th} element
- Median = (31*0.5)+1 = 16.5 = 0.5(16^{th} + 17^{th} elements)

Data presentation

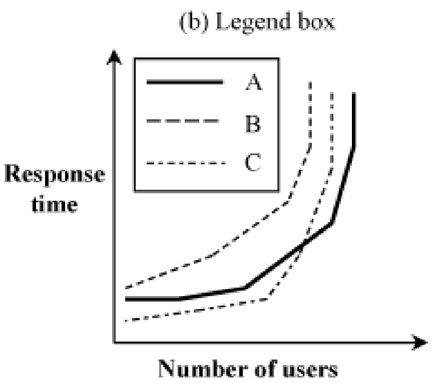
- **One of the important steps in every performance evaluation study is the presentation of final results**
- The eventual aim of every performance analysis is to help in decision making
- An analysis whose results cannot be understood by the decision makers is as good as one that was never performed
- The analysis has to be presented as clearly and simply as possible
- Graphic charts are commonly used in presenting performance results
  - A picture is worth a thousand words
  - A graphic chart saves reader' time and present information concisely
  - Figures allow to quickly grasp the main points of the study and read the text only for details

Type of graphic chart

- The type of graphic chart to be used depends upon the type of variable
- If x is a discrete variable → column or bar chart



- If x is a continuous variable → line chart



Tool

Command-line driven graphing utility

- http://www.gnuplot.info
- http://gnuplot.sourceforge.net/demo/