SPE
Systems – Common Mistakes – PE tools

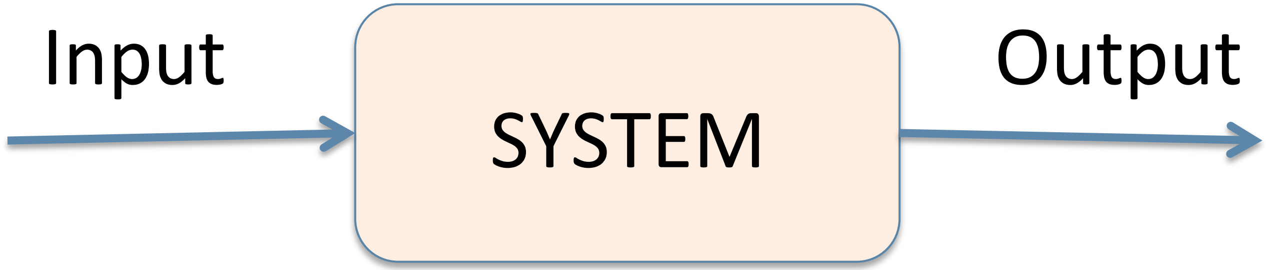
Renato Lo Cigno

http://disi.unitn.it/locigno/index.php/teaching-duties/spe

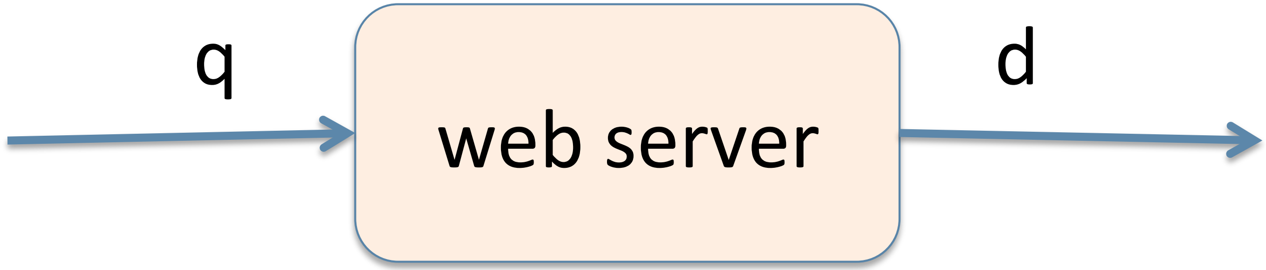
Outline

- What is a system?
- Common mistakes in PE
- PE tools
  - Measures
  - Simulations
  - Analysis

- A "system" is any physical or logical ensemble of which we want to measure some metrics
- Normally we whish to predict some "output" given some "input"



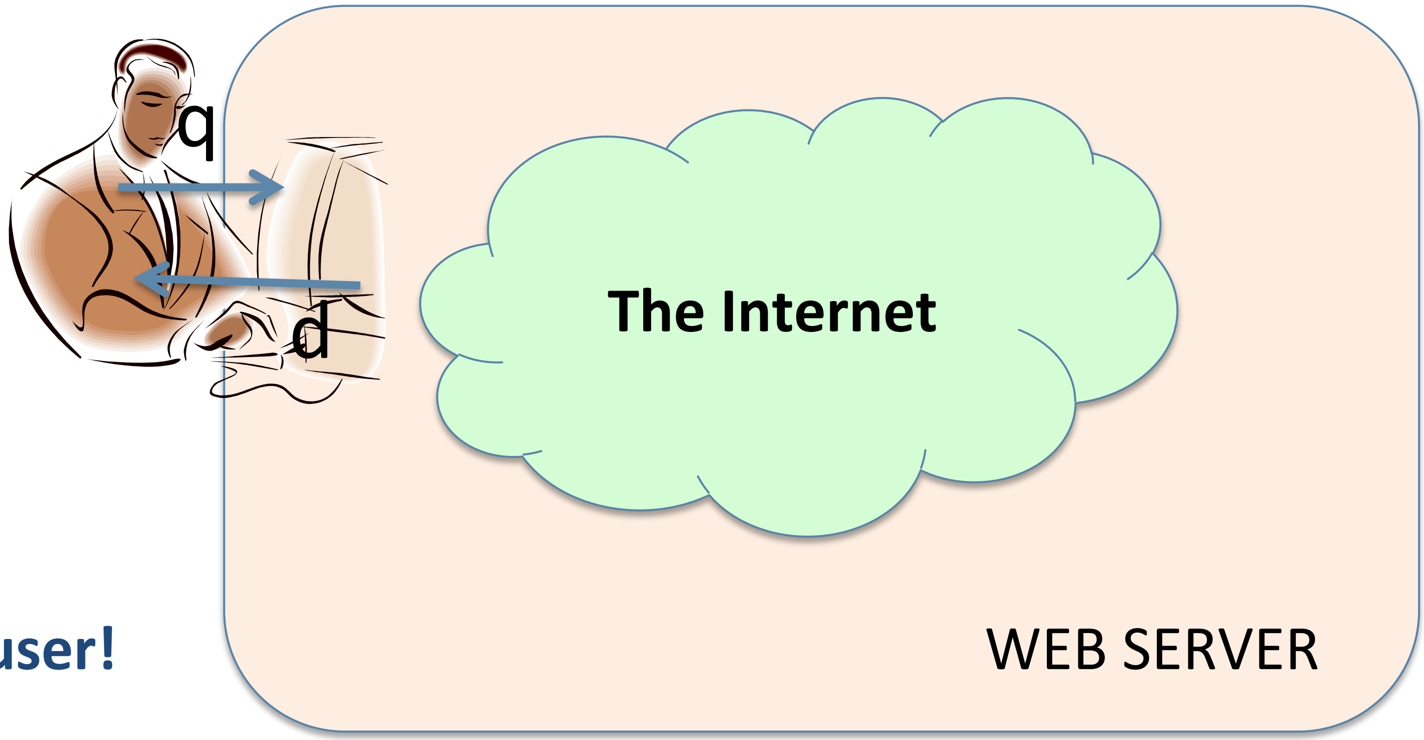
- We are given a "web server"
- Evaluate the service "responsiveness"
- Inputs = queries q_i
- Output = answer delays d_i



Systems: Ex. 1

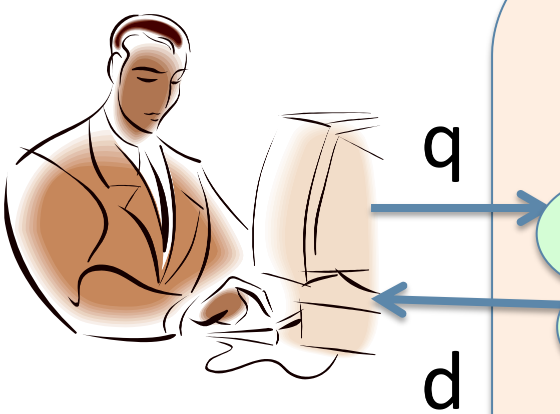
- How do we define a "web server", what is the goal of our evaluation
  - Hardware?
  - Software?
  - What are the "measurement points" of a web server?
- Is a delay a good measure of responsiveness?
- What are the queries q_i that are meaningful for the evaluation?
- What is the "correct" sequence of queries
  - In PE this is normally called the "workload" of the system under evaluation and its choice is of the utmost importance

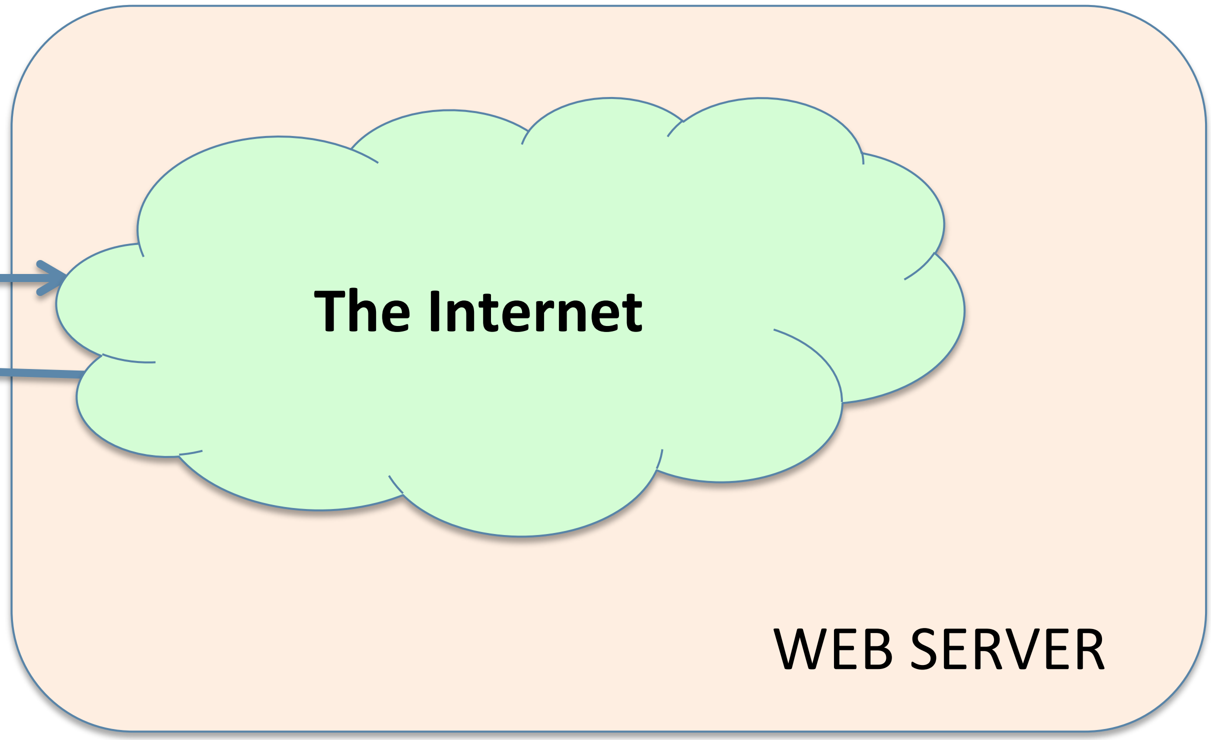
- What is a "web server"?
  - (Answer) or Model 1



good for the user!

- What is a “web server”?
  - Model 1 bis

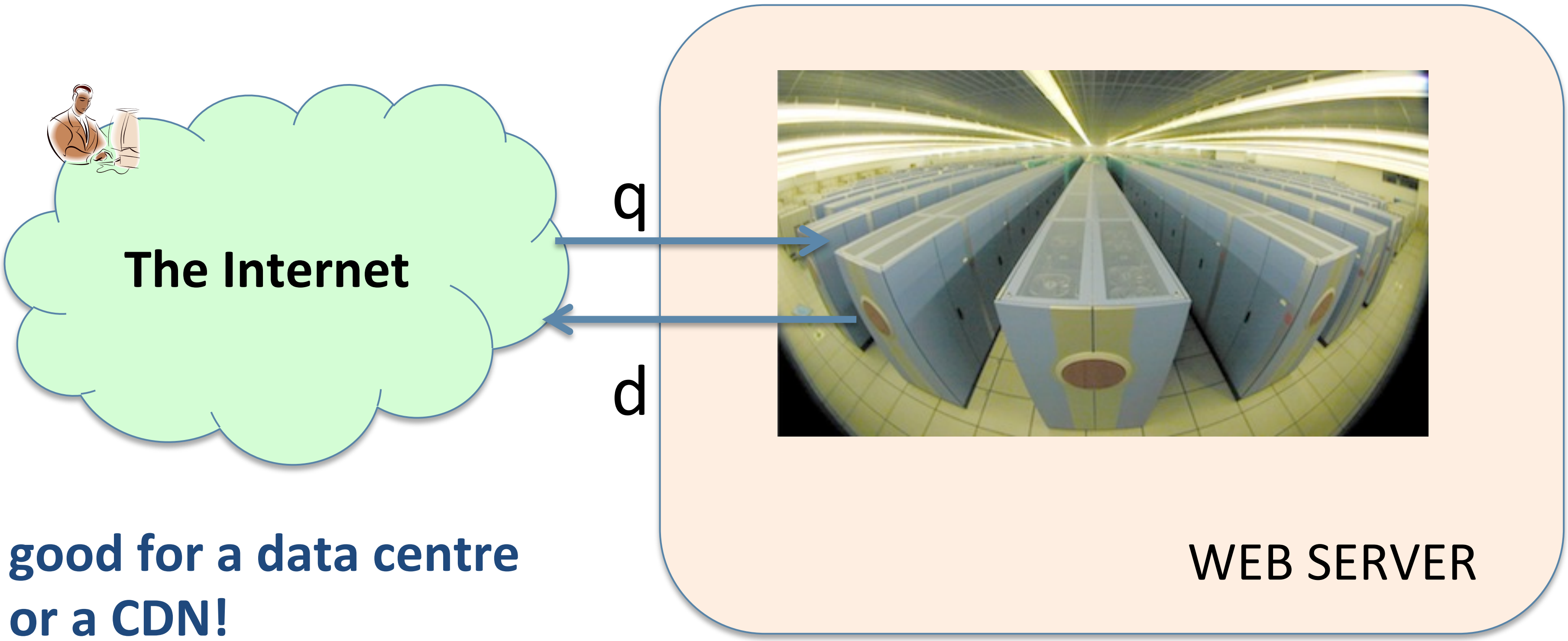




good for the ISP or network manager (your PC is not my business)!

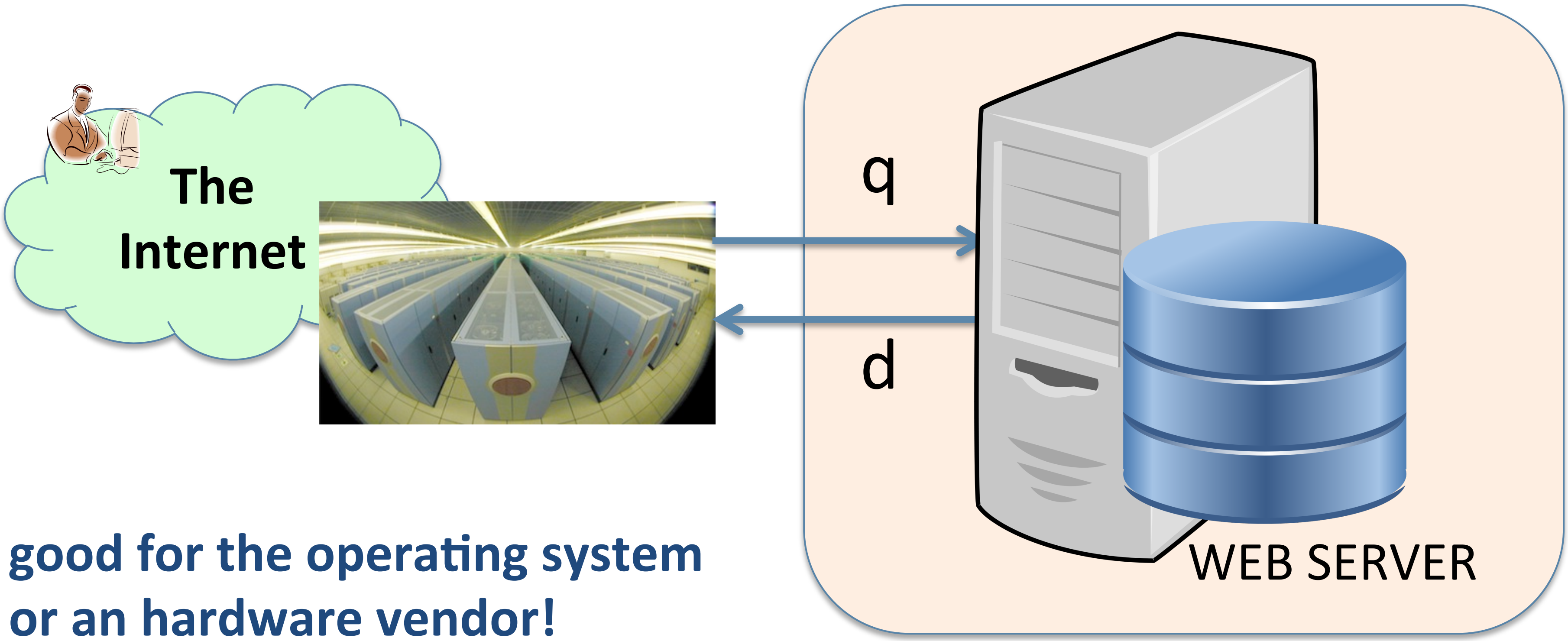
Systems: Ex. 1

- What is a “web server”?
  - Model 2



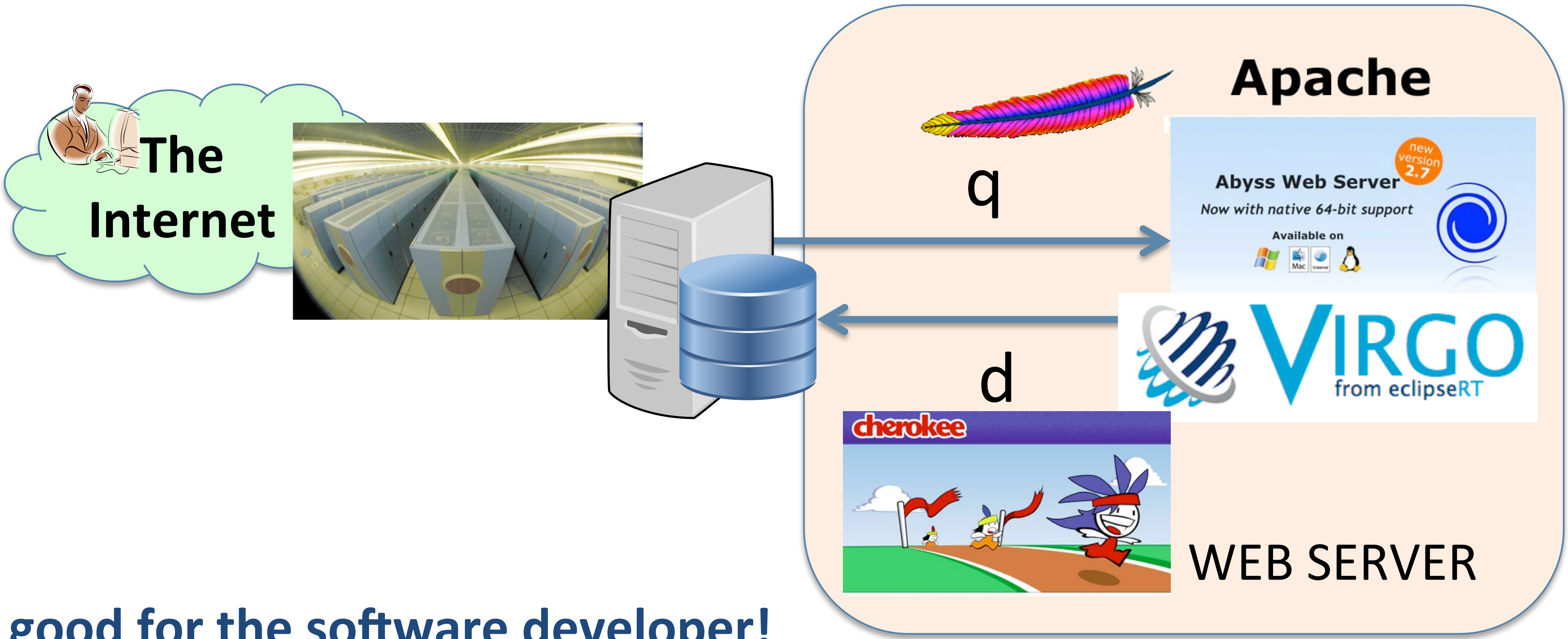
good for a data centre or a CDN!

- What is a “web server”?
  - Model 3



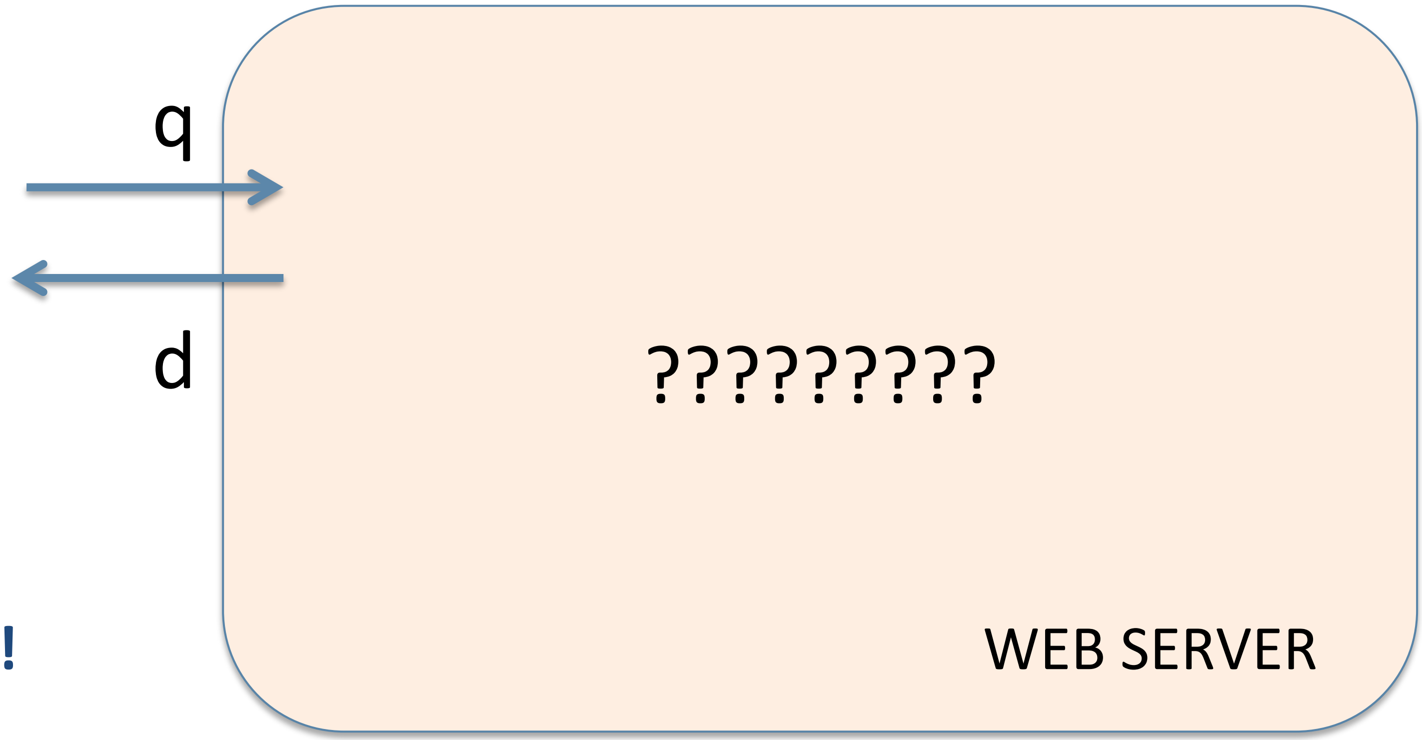
good for the operating system or an hardware vendor!

- What is a “web server”?
  - Model 4



good for the software developer!

- What is a “web server”?
  - Model 5



good for you!!

Systems: Ex. 1

- Different models may require different evaluation tools or techniques
  - Model 1 (specially 1bis) can be easily measured if you seek a “local” answer, but measuring “on average” can simply be impossible
  - Models 2 and 3 can be also easily measured
    - but measures do not give “what-if” answers, so they are not good for design
  - Model 4 is very difficult to measure
- All models can be simulated, but we need to design the simulator correctly
- All models can are prone to analytical solutions, which are fast and easy, but can be very “rough”
  - Very useful to “dimension” systems

- The workload
  - $q_1 \rightarrow d_1$; $q_2 \rightarrow d_2$; $q_3 \rightarrow d_3$; $q_4 \rightarrow d_4$; ...
  - Is it meaningful?
  - Is it representative of all real situations?
- Queries are not isolated and arrive with temporal relationships
  - $Q = \{q_1, t_1; q_2, t_2; q_3, t_3; q_4, t_4; \ldots\}$; q_i in $\mathbf{q}$; t_i in $\mathbf{R}^+$ or $\mathbf{Z}$
  - where $\mathbf{q}$ is the set of all possible queries
  - $\mathbf{R}^+$ and $\mathbf{Z}$ are real or natural numbers (including zero) depending on time being continuous or discrete

Systems: Ex. 1

- The workload
- Does the sequence and temporal distribution of the workload influence results?
  - $Q = \{q_1, t_1; q_2, t_2; q_3, t_3; q_4, t_4; \ldots\} \rightarrow \{d_1; d_2; d_3; d_4; \ldots\}$
  - imply that
  - $Q = \{q_3, t_1; q_2, t_2; q_1, t_3; q_4, t_4; \ldots\} \rightarrow \{d_3; d_2; d_1; d_4; \ldots\}$
  - **or not??**
- The workload is often called also the "**arrival process**" to the system
- Notice that the PE output is normally NOT the output/service of the system
  - A web server does not yield "delays" it returns web pages!!

Systems: Ex. 1

- The output: responsiveness
  - What delay are we measuring?
  - Model 1: is the time needed for transmission and for the browser rendering meaningful?
  - Model 4: Database interactions of a CMS is our business?
  - Model 2: If q_i requires back-end computation (e.g., Hadoop), is that to be included or separated?
  - Model 5: Where do you put the probes to measure in-OS performance?

Common Conceptual Mistakes (CCM)

- Projects often fails due to PE analysis failures
  - The system designed does not perform as it should
  - The software project lags way behind schedule
    - Components have not been analysed
    - The development cycle has not been analysed
  - The nuclear reactor explodes when it should not (Chernobil, Fukushima, Three Mile Islands, Superphoenis, ...)
    - The system reliability has not been properly included in PE (TMI)
    - The system dependability has not been included at all (Fukushima)
- PE Analysis often fails due to initial conceptual mistakes

CCM: Analysis without a Model

- AKA lack of problem understanding
- Without a model we measure a black-box, not a system
- The measures taken may answer "what", but never "how" and "why"
- Artificial Intelligence & Machine Learning try to do some analysis on black-boxes
  - They have sophisticated stochastic models behind them
  - The results remains valid until the stochastic model behind is valid
    - e.g., the model assumes that the output distribution has a single mode, if not true the results are not valid

CCM: Search for the "True" Model

- ***"All Models are Wrong ... some of them are Useful"***
  (Mark Tweine, Albert Einstain, George Box or somebody else much more famous than me)

- **Alternative version 1: *"All Models are Right ... most of them are Useless"***
  (Unknown, reported by T. Tarpey in a seminar in 2009)

- **Alternative version 2: *"All Predictions are correct ... Unless they regard the Future"***
  (maybe Mark Twaine, if not, myself)

CCM: Biased Goals

- Frequent mistakes in Science & Engineering
  - I want to show that my system is better than his one
  - My boss will fire me if I show that his idea is bogus
- Always present in brochures & commercial datasheets
  - Those are ads, not science
- More often simply lack of pre-analysis
  - I have not given my problem enough time, so my global understanding is not enough to drive the PE process
- Pre-concepts prevent seeing the truth
  - I expect a red ball, and never see the green cube popping out

CCM: No Goals & Random Search

- I don't know what I need to measure/model
  - Measures taken cannot be interpreted
  - The model I make finally yields no meaningful outputs
    - e.g., results are always constant even changing parameters
- I have a model, but change parameters at random
  - Results taken in this way do not offer insights
- I have so many variables and parameters that the search space is practically infinite
  - Any amount of results will always be negligible w.r.t. to space to be analysed
  - Typical mistake in simulations

- Selection of the correct tool (measures, simulations, analysis) is fundamental
  - e.g., we cannot "measure" a disaster, it's too late!
  - More later on
- The model must be tailored for the problem
  - Too much details makes it cumbersome & awkward to use
  - Too few details makes results unreliable

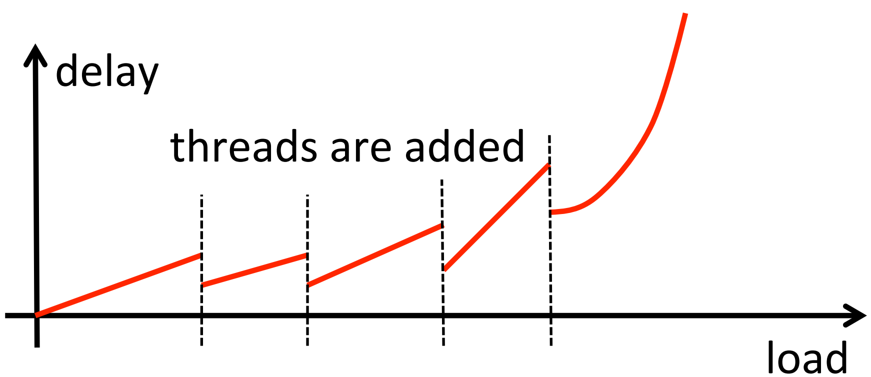
CCM: Uncheked Workload

- Outputs depends on the inputs
  - A different input than "thought" leads to wrong interpretations
- Biased workload
  - a skewer random generator
  - a workload that include rare events that never happen during the simulation/measure ... yet they are possible
- Changing one parameter of the workload (e.g., adding new web queries) changes also other parameters that are not controlled (e.g., interaction with data-bases)
  - Also changing variability I also change the mean ...

CCM: Impossibles & Outliers

- Outliers (results that are discarded as not meaningful) must always be checked
  - Instrument fault?
  - “Unlucky” Simulation?
  - Or a mistake in the model/implementation?
- Events that should not happen **must not** be ignored
  - Symptoms of a different system behaviour in measures
  - Symptoms of “eisenbugs” in software (simulations)
  - Symptoms of numerical problems in analytical solutions
  - ...

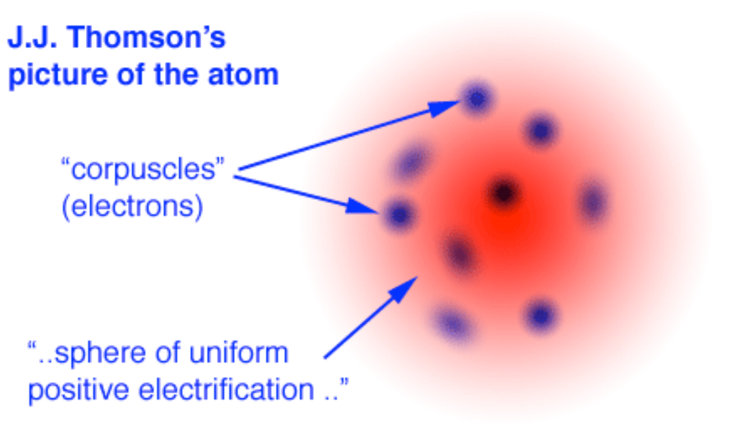
- This is a “variation” of Impossibles & Outliers
- When building the model some important factor is left out
  - Result change when they should not (the “factor” has changed, but we ignore it)
  - We trust our results ... but they are wrong
- E.g., we model a web-server as a single process, while indeed it is multi-threaded with threads added based on a performance threshold



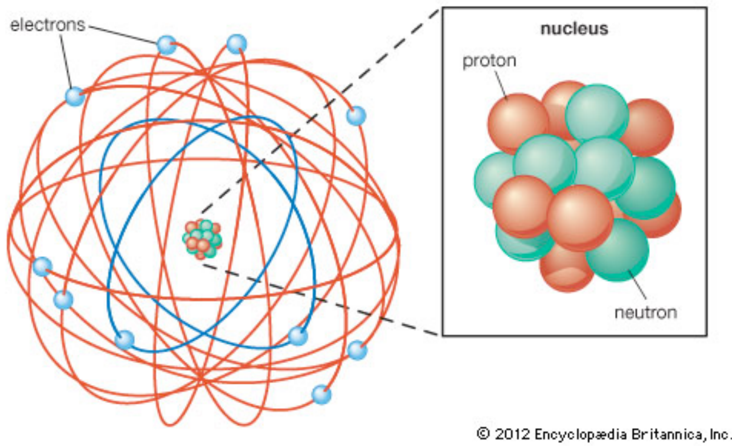
- If two models give the same output the simpler is preferred
- If a model has more than 100 parameters you can obtain any result you want ...
  - This becomes PE as “Parameters Engineering” not as “Performance Evaluation”
- An esoteric explanation is often appealing, but not necessarily the the correct one

- The model is never put in discussion ... but sometimes it is wrong
- Rutherford vs Thomson Atom is the perfect example
  - The interpretation of scattering was simply impossible with Thomson model

Thomson model



Rutherford model

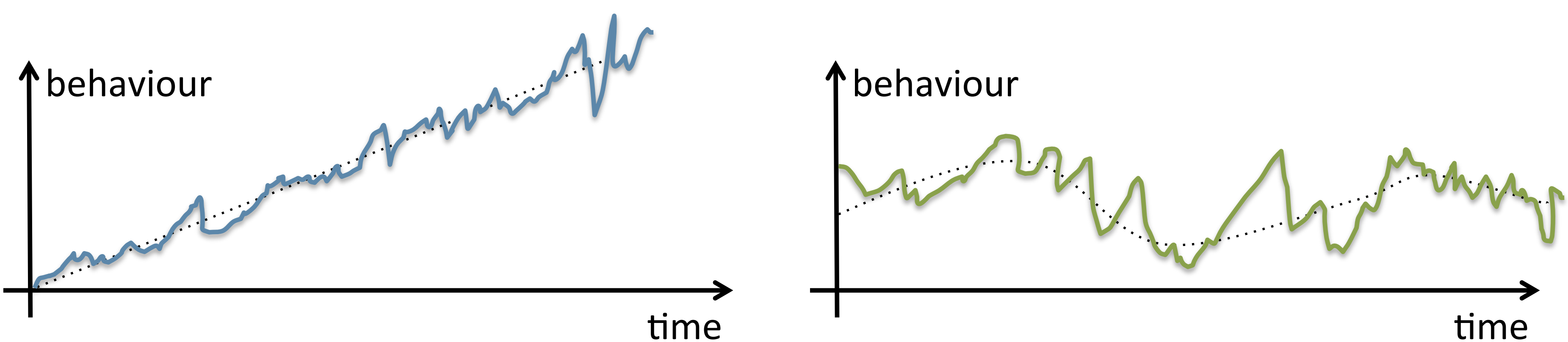


CCM: Stationarity Assumption

- Stationarity is the property of not changing the behaviour in time:

  $$G(t) \approx G(t-t_0) \text{ for all } t_0$$

  where ≈ indicates a stochastic equivalent behaviour
- Not all systems and phenomena are stationary
  - Interpreting non-stationary systems in light of a stationary model is wrong



CCM: Assumptions & Presentations

- Assumptions and approximations must be very clear (and declared!!)
  - A PE exercise can be OK if I know its goals and validity, but can be rejected entirely if I think it is about the real system
  - E.g., a multi-core CPU approximated as a single core one
- Results presentation is fundamental
  - Graphs are better than tables & numbers
  - Results must be presented with their confidence
  - The results selected for presentation must:
    - Give insight
    - Enable decisions

Measures

- The most traditional "scientific experiment"
- The system must be available, accessible and observable
- Workload generation is fundamental and often difficult
  - E.g., loading a 100Gbit/s optical network
- Changing system parameters can be difficult (impossible sometimes)
- Setting probes to take measures might be difficult (impossible sometimes)
  - E.g., adding a hardware counter to a CPU
- Measures must be repeated to achieve confidence

Measures

- They are often considered the most reliable PE form
  - But they do not give predictions!
  - You cannot “measure” disasters, you must avoid them!!
- Measures are intrinsically affected by noise
  - Can be additive, multiplicative, or even distortive
  - In any case it is represented by RVs
- Metrics themselves can be stochastic in nature
  - E.g., replication speed of cells; equilibriums in chemicals at high temperature; completion times of jobs in loaded systems; delays in the Internet; ...

- A software that mimics the behaviour of the system
- Requires and abstract model
- State parameters of the system must be carefully selected

$$S(t) = \{s_1, s_2, \ldots, s_N\}$$

  - Parameter ranges are important and must be checked
  - t is inherently discrete: t in $\mathbb{Z}$
- The evolution of the system is described as conditional random variables

$$S(t+1) = V|S(t)$$

where V is a vector of stationary R.V.

Simulations

- Workload generation is easier than in measures
- Parametric studies are normally easy
  - But they have a large computational cost
- A simulation “run” is a random walk in the state space of the model
- Simulation runs can be interpreted as Monte-Carlo solutions of the set of couples differential equations that describe the state evolution
- Multiple runs (or batch means techniques) are needed to evaluate the confidence of results

Analysis

- Finding the model is difficult
- Finding its solutions sometimes even more
- Monte-Carlo numerical integration is normally feasible
  - In this case the analysis becomes hybrid with simulations, but the confidence is normally easier to estimate and sometimes it does not require multiple solutions
- Parametric studies are very simple and effective
- Analytical models are often in the form of dynamic equations
  - i.e., coupled differential equations

Analysis: Example of a model

$$\begin{cases} dG(t)/dt = aF(t) & G(0) = G_0 \\ dF(t)/dt = -bG(t) & F(0) = F_0 \end{cases}$$

- A set of coupled differential equations
- This example can describe the evolution of a population G as a function of resources availability F, which in turn decreases as the population G increases
  - Is there an equilibrium?
  - Will G survive or get extinct

Analysis

- Finding the model is difficult
- Finding its solutions sometimes even more
- Monte-Carlo numerical integration is normally feasible
  – In this case the analysis becomes hybrid with simulations, but the confidence is normally easier to estimate and sometimes it does not require multiple solutions
- Parametric studies are very simple and effective
- Analytical models are often in the form of dynamic equations
  – i.e., coupled differential equations

- Measures confirmed with simulations or analysis
  - Insight & Results interpretation
- Analysis confirmed by simulations
  - Be careful not to use the same model!
- Analysis confirmed & tuned with measures
- Simulations confirmed and tuned with measures
- Simulations confirmed by analysis

- Never trust a single tool for PE (if possible)

Tool selection & Tradeoff

Criterion/Metric	Measure	Simulation	Analysis
Stage of design	Prototype – Operational	Any	Any – Very Early
Time required	Variable – Long	Variable	Small
Design complexity	High	Low – Medium	Variable
Accuracy	Variable	Variable	Variable
What-If evaluation	Difficult	Computationally expensive	Easy
Cost	Very High	Low to Medium	Low
Interpretation	Difficult	Medium	Easy
Marketing	Easy	Medium	Difficult (lack of understanding)