Introduction to Networking and Systems Measurements

Lecture 1: Introduction

Dr Noa Zilberman
noa.zilberman@cl.cam.ac.uk

Prof Andrew W. Moore
andrew.moore@cl.cam.ac.uk

Administrivia

Scope:

- Characterization and modelling of systems and networks using measurements.

Course structure:

- Lectures – 6 hours – (1st one FS09 2nd-5th FS07
- Guided Labs – 10 hours - SW02 (ACS lab)

Assessment:

- Lab writeups (20%) - 18/11/2019 12:00
- Evaluation of an artefact – 5000 words paper (80%) – 6/12/2019 12:00

Schedule

Week	Lecture	Lab
1	- Introduction to Performance Measurements - Basic Measurements	
2	Advanced Measurements	Basic Measurements
3	Reproducible Experiments	Traffic Capture & Latency
4	Measurements Pitfalls	Traffic Generation
5	Device and System Characterization	Study of an Artifact
6		Reproducibility

Some logistics for 2018-19

Web page: http://www.cl.cam.ac.uk/teaching/current/L50/

Repository: https://github.com/cucl-srg/L50

Mailing list: *cl-acs-l50-announce@cam.ac.uk*

Grades:

Mphil (ACS) – Pass / Fail - based on a mark out of 100
All others (DTC) – Mark out of 100

Next steps

- Explore the web page and repo

  http://www.cl.cam.ac.uk/teaching/current/L50/

  https://github.com/cucl-srg/L50

- Decide if you still want to take the class - promptly

Why Measurements?

We measure things every day, all the time:

- How *far* is the destination? - distance
- How *long* will it take to travel? – time
- How *much* will it cost? – price

We also measure CS-related aspects:

- How *fast* is the CPU? – frequency
- How *big* is the file? – storage size
- How *much* power is used? – power

System Measurements

Can be used to answer questions such as:

- Is this system working as expected?
- Is this system better that another system?
- What are the limitations of my system?
- Where are the system's bottleneck?

Network Measurements

Can be used to answer questions such as:

- What is the topology of the network?
- Are there performance issues?
- What are the network’s bottlenecks?
- How do devices that connect to the network operate?

Performance – not just bits per second

Second order effects

- Image/Audio quality

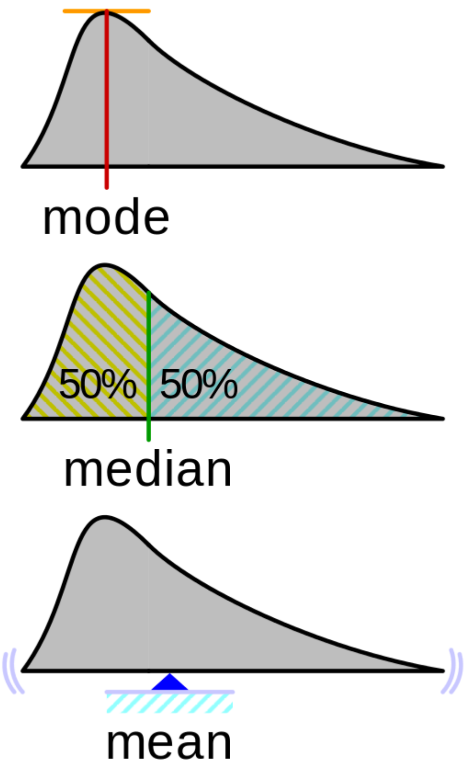
Other metrics…

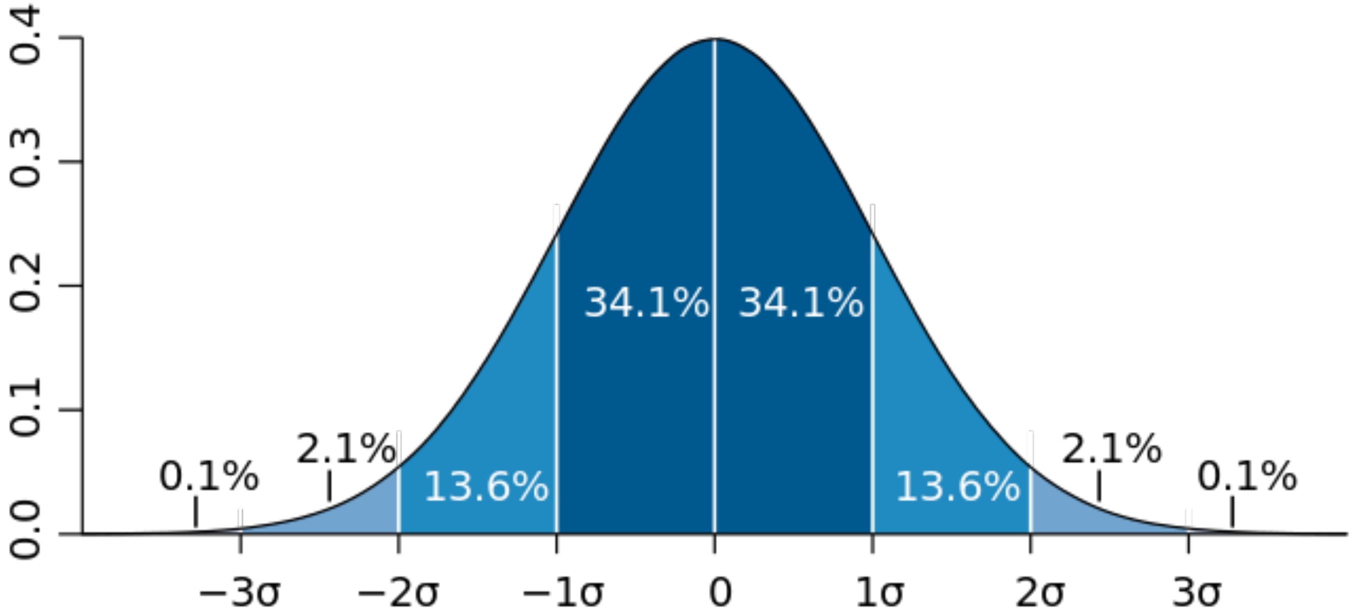
- Network efficiency (good-put *versus* throughput)
- User Experience? (World Wide Wait)
- Network connectivity expectations
- Others?

Statistics in Measurements
Terms and limits

- Mean
- Median
- Standard deviation
- Independence
- Heavy tail distribution (and where it all goes wrong)
- Probability density function / Histogram

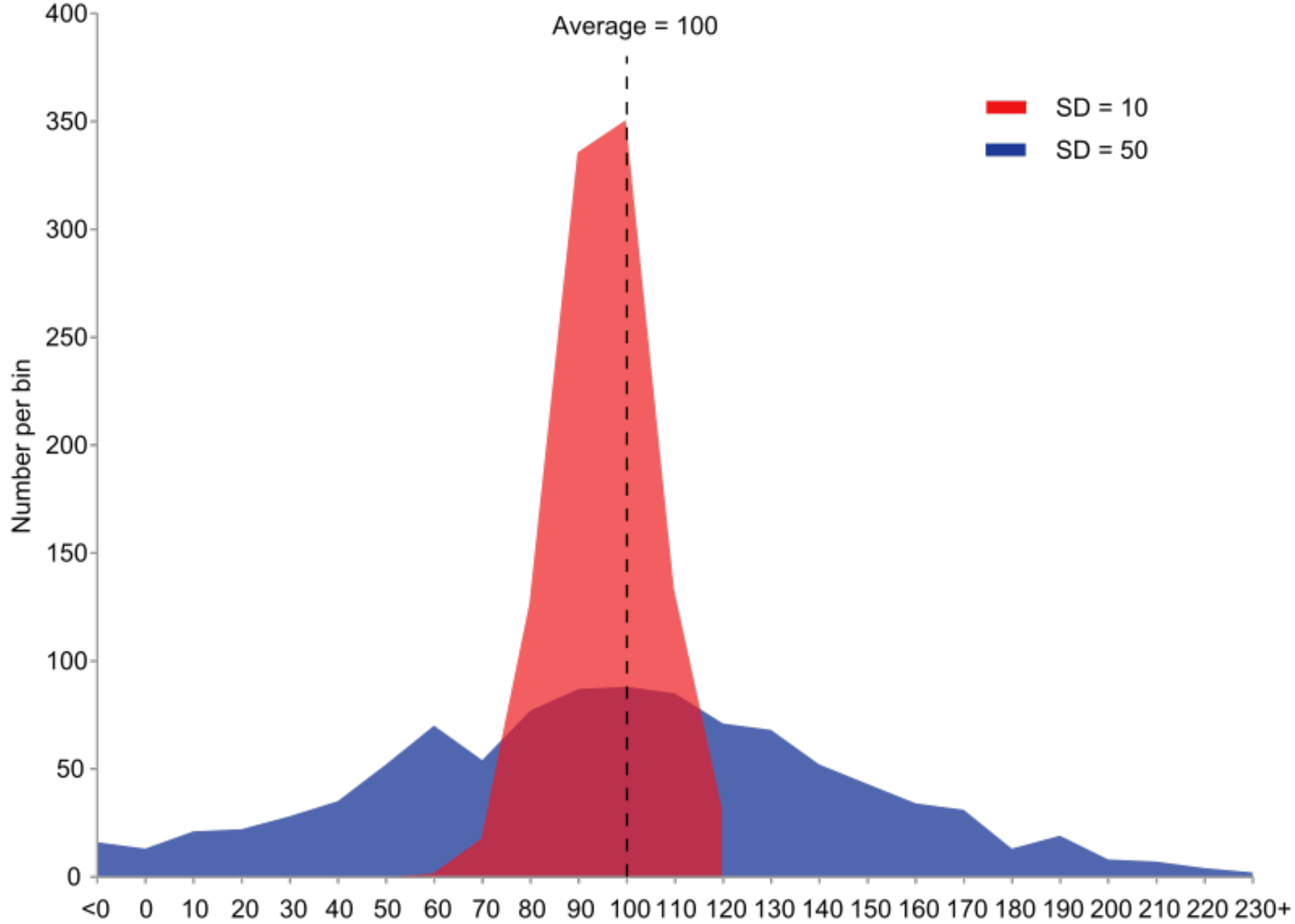
  Cumulative density function (CDF) and CCDF
- Tests (two variable or hypothesis: t-test, multivariable: ANOVA)





Standard Deviation in a Normal Distribution

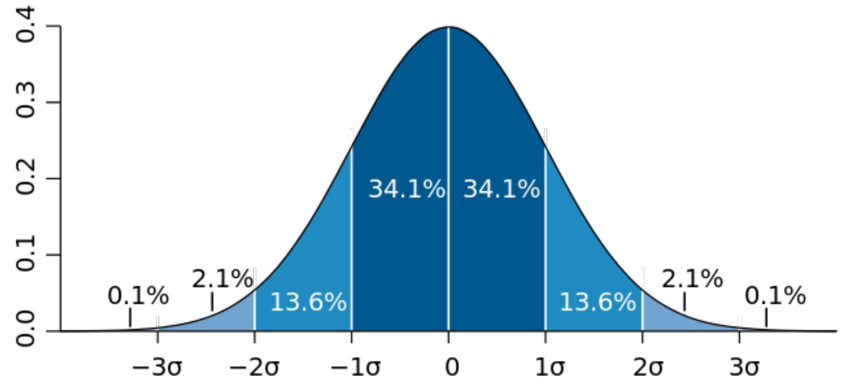
Two sets of samples with the same mean and different Standard Deviations



Confidence Intervals? Error Bars? Sample Size?

- **Confidence Interval** is the interval (range) of values you have confidence a given sample will fall within….

- **Error Bar** represents the range of all values for an experiment (sometimes the confidence interval is used – this makes assumptions!)



- **Sample Size** is the number of (measurements) made

    certain *tests* (eg t-test) can assist us in deciding on a sample size

    when we don't choose the sample size those same tests will declare

    the confidence to hold in how representative the sample-set was

Why our most-basic assumptions are wrong

or Why Independence is not a great assumption…

We measure the use of electricity in a neighbourhood over a day

There is a popular TV programme

A commercial break sees much of the population in the neighbourhood *putting the kettle on*

This is a correlated event (not independent)

Correlation is also a common phenomena in the Internet

At many timescales (weekly, daily, hourly, predictable functions of time, distance, computer-type, application-type, favourite soda….)

Why our most-basic assumptions are wrong

Independence – why we care

- Some(many/most) analysis techniques assume independence
  - Highly correlated events may mean ***non-representative*** measurements
- We might use measured data as-if it was independent/representative

What can we do?

- Constant vigilance:
- Look at the data, best-practice, *think*.

Why our most-basic assumptions are wrong

- Why Poisson distribution is not a great assumption…

We measure the use of electricity of 1000 households to determine average use as a representation (informed guess) for the nation….

Households have a high prevalence of solar panels

Not so presentative…..

This example might give a skewed distribution

This is only one cause of normal distribution failure

Distributions

- Normal Poisson Binomial….. Not the same and often ‘jumbled up’

- A **Normal distribution** is continuous
- A **Poisson distribution** is discrete
- A **Poisson** random variable is always $[0,\infty)$

- It is common to mean Poisson even if people say Normal….

Why our most-basic assumptions are wrong

Poisson distributions– why we care

- Poisson distributions make analysis and interpretation easy
  (e.g. mean, standard deviation, etc.)

What can we do?

- Look at the data, best-practice, *think*.
  - Particularly when the dataset is small

- Did I mention that normal distributions assume independence? <sad face>

Central Limit Theorem or “Mix enough to get Normal”

- CLT says that statistics computed from the mean (eg the mean itself) are approximately normally distributed – regardless of the distribution of the population

(OR ANOTHER WAY)

- CLT says the more data you have the more the observed mean will become the true mean

- Sadly CLT can say nothing about variance!

Law of Large Numbers or “You just need more data”

- LLN is actually a handy idea that says ”given enough data and obey the rules, the sample (measurements/overvations) will better represent the population (causal) characteristics”

- Sadly the rules are
  - Independence (again)
  - Population should not be skewed (eg be larger than *30, or is it 40? 400?....)*

- LLN is useful, it tells us lots of things:
  - <if rules> - the average of more data observations becomes the mean of the source of observations
  - But LLN says nothing about the variance.

When Standard Deviations go wrong…

- Standard Deviations (SD) indicate the *dispersion* of the underlying data

but SD measures build in some assumptions: symmetry and common computation assume a Poisson distribution….

Sometimes they simply don’t capture the nature of the data, nothing showed this up more clearly than the heavy-tail distribution…..

Heavy Tails… (condensing a lot into a slide)

- Certain phenomena (eg correlated events) can cause unusual (rare) events
- These events led to very large (wide) distributions, ones where the tail(s) has more values than a Poisson distribution would predict
- The more dispersed the data : the larger the Standard Deviation measure
- One definition of heavy tails is where Standard Deviation tends to infinity….
- Sadly, heavy tail distributions are very (VERY) common

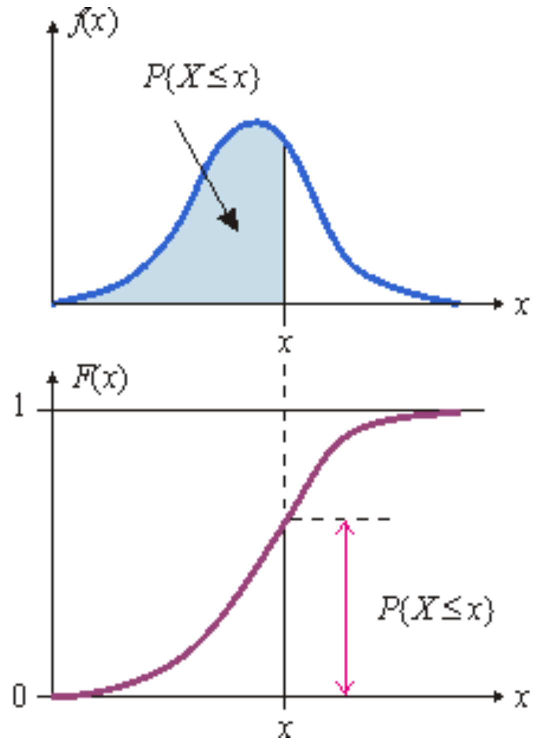
"1 in a million events occur about 9 times out of ten" – T. Pratchett

How to read a PDF CDF and CCDF…..

- A Probability Density Function tells me the probability for a specific value

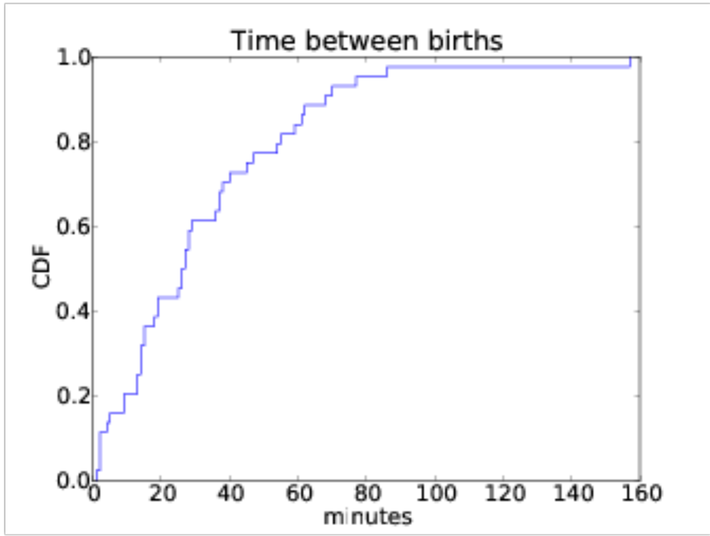
- A Cumulative Density Function is a sum of probabilities

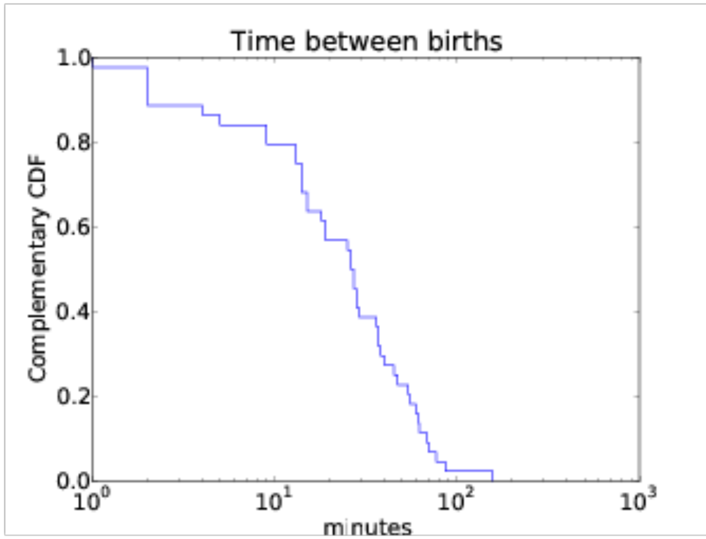
That is: "is the probability that the random variable will take a value less than or equal to a particular level."



How to read a (C)CDF…..

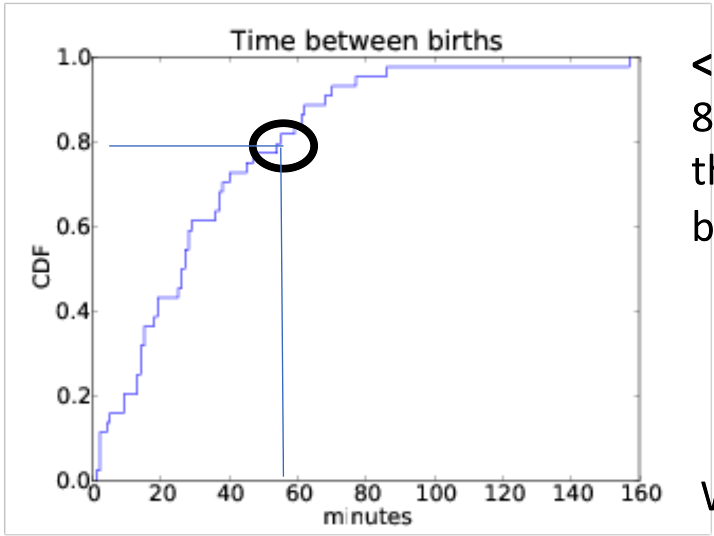
- A Complementary Cumulative Density Function 1-the sum of probabilities
  - Useful for "how often the random variable is(at or) *above* a particular level."





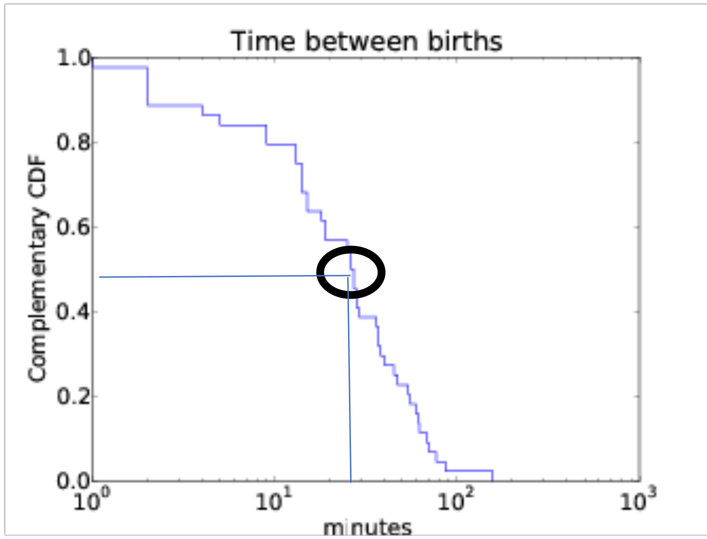
How to read a (C)CDF…..

- A Complementary Cumulative Density Function 1-the sum of probabilities
  - Useful for "how often the random variable is(at or) *above* a particular level."



<- CDF
80% of the time it was less than 55 minutes between births

CCDF ->
Over half the *time between births* Were longer than 20 minutes



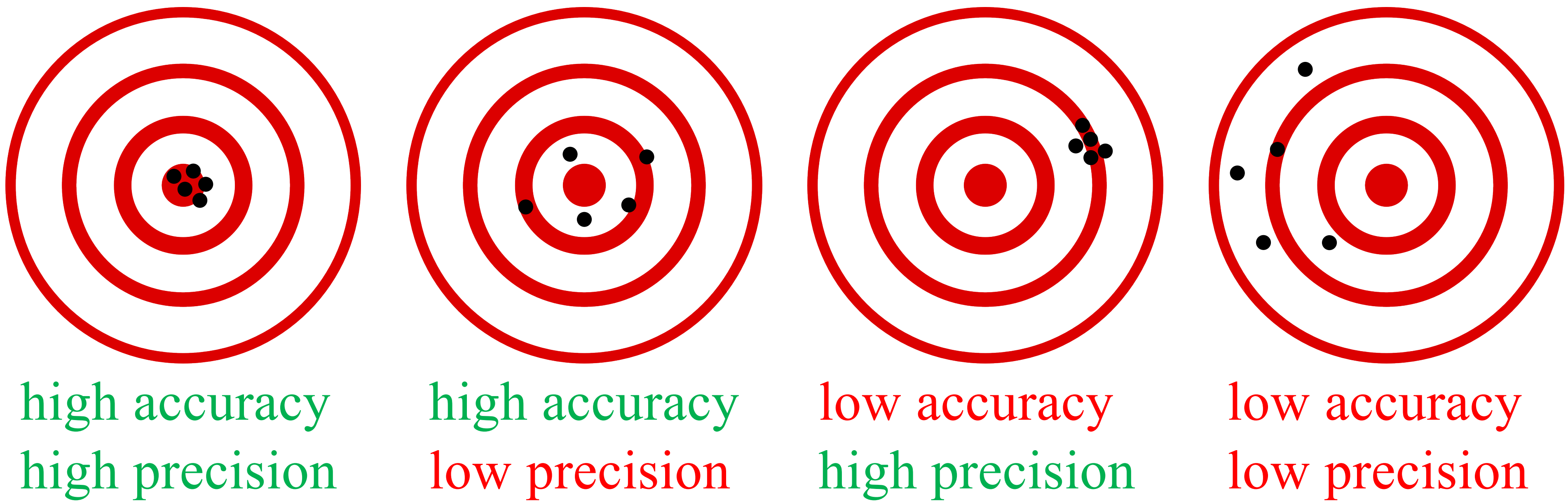
Terminology Matters!

… in greater depth in following weeks

Precision, Accuracy and Resolution

Accuracy – How close is the measured value to the real value?

Precision – How variable are the results?



Precision, Accuracy and Resolution

Resolution – The smallest measurable interval.

The resolution sets an upper limit on the precision.

high resolution

low resolution

In our experiments, resolution many times be determined by clock frequency (directly or indirectly)

Bandwidth, Throughput and Goodput

- Bandwidth – how much data can pass through a channel.
- Throughput – how much data actually travels through a channel.
- Goodput is often referred to as application level throughput.

But bandwidth can be limited below link's capacity and vary over time, throughput can be measured differently from bandwidth etc…..

Speed and Bandwidth

- Higher bandwidth does not necessarily mean higher speed
- E.g. can mean the aggregation of links
  - 100G = 2x50G or 4x25G or 10x10G
  - A very common practice in interconnects

RTT, Latency and FCT

Measures of time:

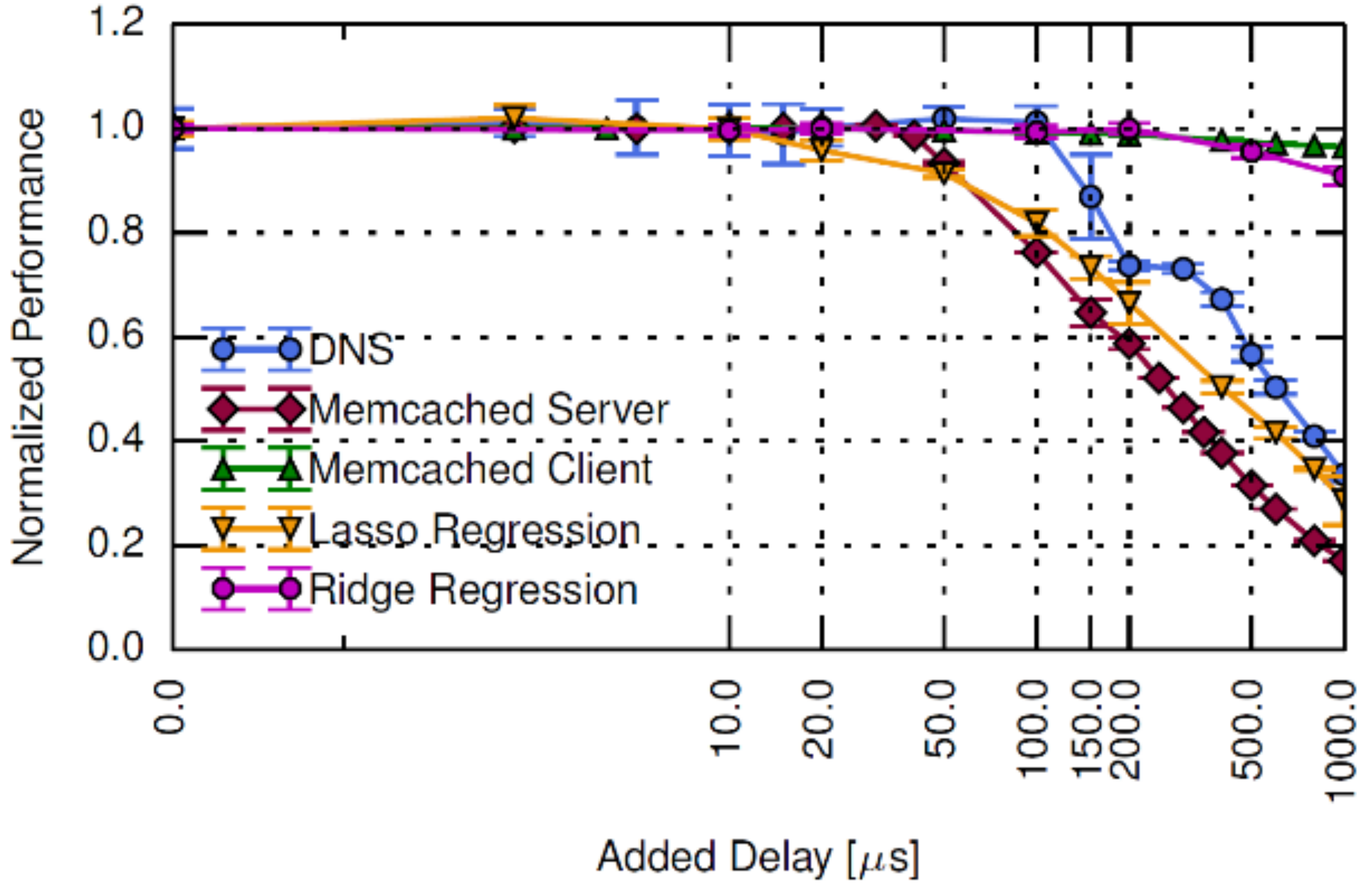
- Latency – The time interval between two events.
- Round Trip Time (RTT) – The time interval between a signal being transmitted and a reply is being received.
- Flow Completion Time (FCT) – The lifetime of a flow.

Performance Metrics

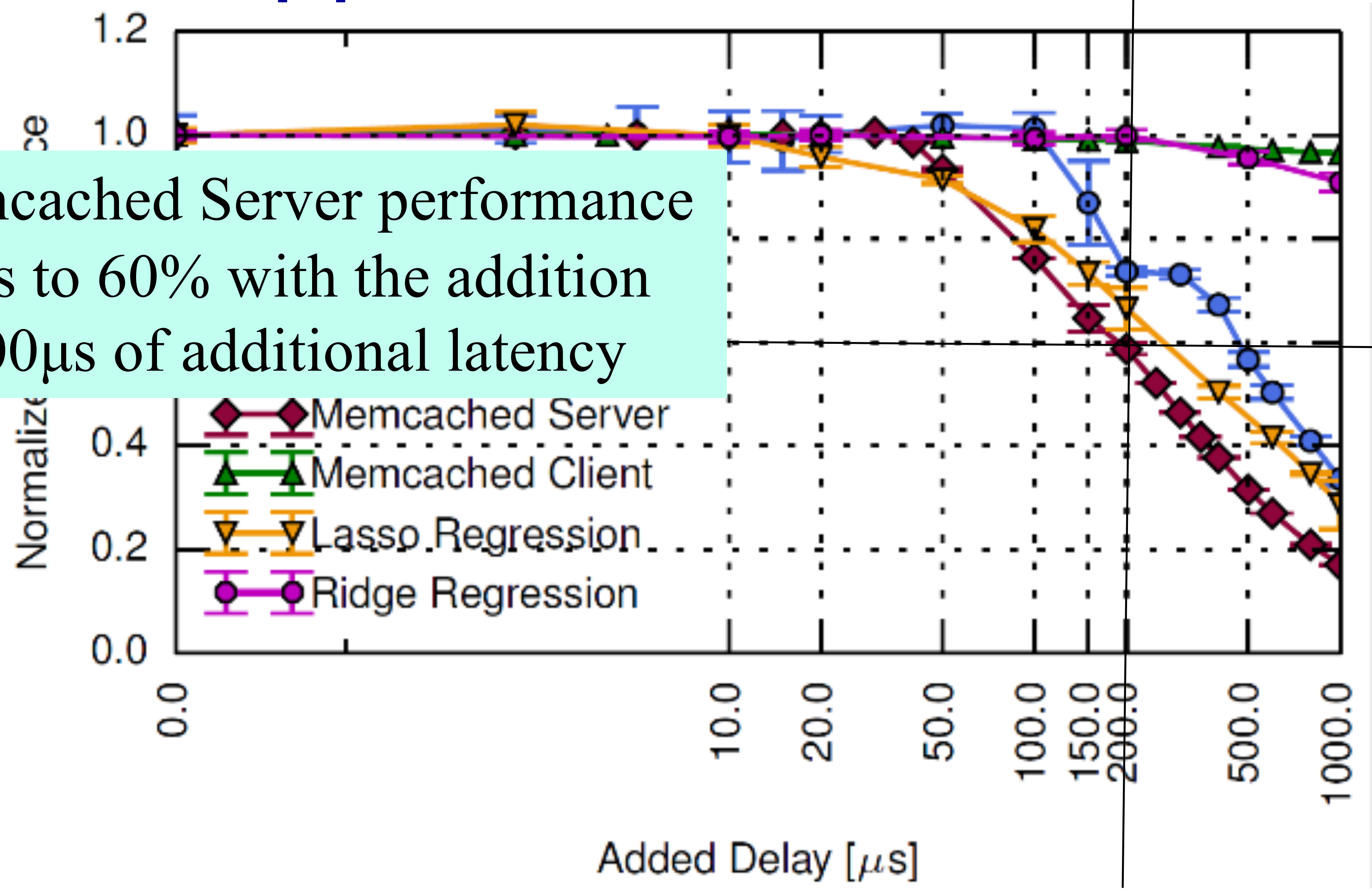
- Throughput, FCT etc. are measures of *Performance.*

- Bandwidth, RTT, packet loss etc. don’t indicate (directly) how good or bad the application / system / network perform.

Example: The Effect of Latency on Application's Performance



Example: The Effect of Latency on Application's Performance



Memcached Server performance drops to 60% with the addition of 200μs of additional latency

Types of Measurements

Measurement Techniques

- Active
  - Issue probe, Analyse response
- Passive
  - Observe events

Example: Active vs. Passive RTT Measurement

- Active measurement – `ping`
  - Sends ICMP Echo Request message
  - Waits for Echo Reply message
  - RTT is the time gap between the request and the reply.
- Passive measurement – `tcptrace`
  - Uses TCP dump files
  - Calculates RTT according to timestamps logged in the dump.

Comparison

Passive	Active
Can only measure in the presence of activity / traffic	Measures even when tapping activity / traffic is not possible
Measures user experience, behaviour Measures protocol exchanges	Measures system, network, application performance
Raise privacy concerns	Adds probing load: - Overload system/network - May bias inferences

Measurement Vantage Point

- Point where measurement host connects to system / network
- Observations often depend on vantage point
  - Do you have enough vantage points?
  - How are the vantage points distributed?
- Can affect, e.g.:
  - Topology discovery
  - Bandwidth analysis

Possible Vantage Points

- End-hosts
  - Active measurements of end-to-end paths
  - Passive measurements of host’s traffic
- Routers/Measurement hosts in network
  - Active measurements of network paths
  - Passive measurements of traffic, protocol exchanges, configuration