

Advanced Systems Lab

Tutorial V

Analyzing a system

G. Alonso

Systems Group

<http://www.systems.ethz.ch>

Analytical treatment of a system

- Find ways to predict the behavior of a system through a model
 - Know what will happen without having to run an experiment
 - Use it to predict behavior
 - Gain more understanding about the system
 - Extrapolate behavior and results beyond what is feasible to test empirically
- Use experimental data to refine and validate the model

Tools

- Statistics
- Queuing theory and models
- Operational laws
- Intuition and insight on how the system works

Example 1

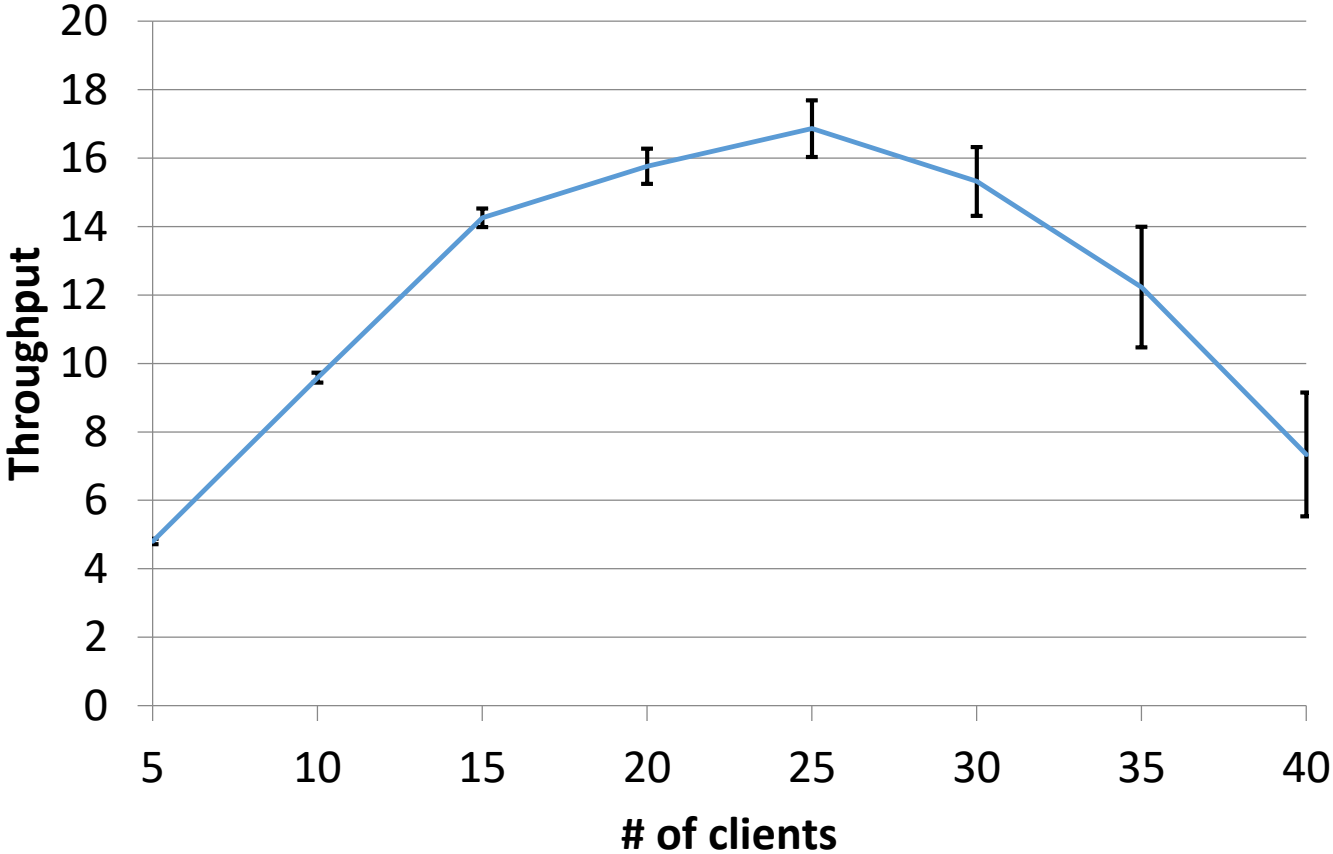
- For each load level (number of clients), five experiments are run and the table shows the throughput for each experiment in operations per second, the average, and the standard deviation. Clients have a think time of 1 second and the experiments are run as a closed system.

Experimental data

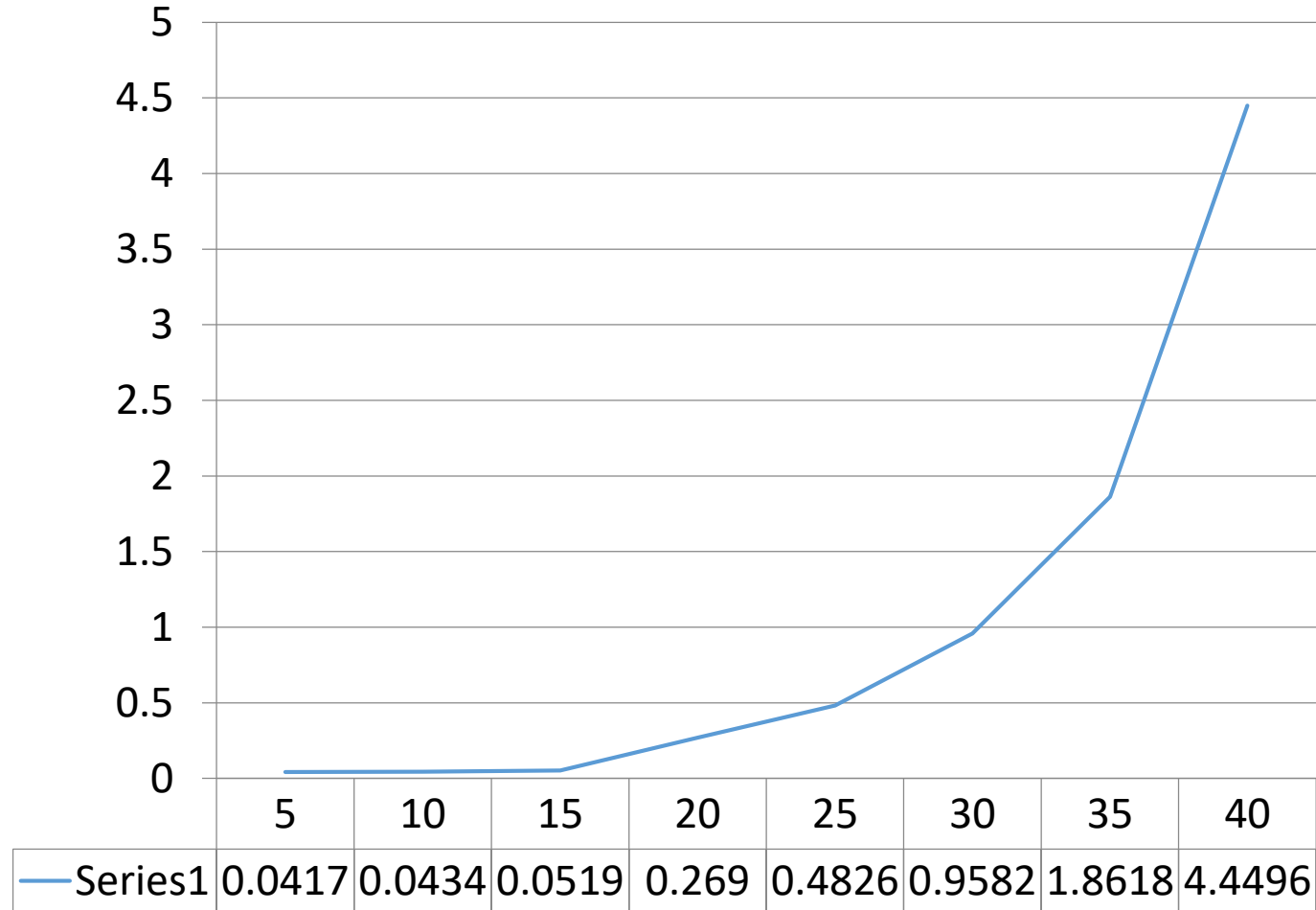
Throughput *Repetition*

Load	1	2	3	4	5	AVG	STD
5	4,60	4,90	5,00	4,70	4,80	4,80	0,16
10	9,40	9,60	9,20	9,92	9,80	9,58	0,29
15	14,20	13,40	14,90	14,40	14,40	14,26	0,55
20	16,10	15,30	17,40	14,80	15,20	15,76	1,03
25	17,20	14,90	17,01	15,90	19,30	16,86	1,65
30	16,50	17,00	12,80	16,80	13,50	15,32	2,00
35	8,30	14,10	10,25	11,20	17,30	12,23	3,52
40	12,50	4,30	5,20	4,90	9,80	7,34	3,62

Throughput (just plot it)



Response time (from Interactive Law)



Pay attention to what the numbers say

LOAD	X	RT
5	4,80	0,041667
10	9,58	0,043406
15	14,26	0,051893
20	15,76	0,269036

At 20 clients the response time is 5 times higher (system is 5 times slower). One cannot reasonably claim that the system supports more than 15 clients or that the maximum throughput is ~17 jobs per second

Service rate

- Based on those numbers:
 - What would be the service rate of the system (modeled as an M/M/1)
 - How would you calculate the number of jobs in the system?
 - 5 0,2
 - 10 0,416
 - 15 0,74
 - 20 4,24
 - 25 8,138

And now what?

- You have characterized the system
 - Maximum performance
 - Inflexion points
 - Service rates at different loads
 - Queue length
 - ...
- Can you explain why these things are the way they are? => map to the design

Further analysis

- Start with system as one block, then divide in smaller parts and see what you can say about each one of them
 - Where is the bottleneck?
 - In the middleware, where is the bottleneck?
- Use 2k experiments approach to evaluate more aspects of the system once you know where the bottlenecks are