IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2018

MEng Honours Degree in Mathematics and Computer Science Part IV
MEng Honours Degrees in Computing Part IV
MSc in Advanced Computing
MSc in Computing Science (Specialist)
for Internal Students of the Imperial College of Science, Technology and Medicine

*This paper is also taken for the relevant examinations for the
Associateship of the City and Guilds of London Institute*

PAPER C436H

PERFORMANCE ENGINEERING

Thursday 22 March 2018, 11:40
Duration: 70 minutes

Answer TWO questions

Paper contains 3 questions
Calculators not required

1a   A basic e-commerce website consists of the following web pages: *Home* (H), *Add* (A), *Buy* (B), and *Catalog* (C). The following user navigation sessions have been recorded in the web server log files:

- $H \rightarrow C$
- $H \rightarrow C \rightarrow A \rightarrow B$
- $H \rightarrow C \rightarrow H$

i)   Draw a user behaviour graph (UBG) that models the observed sessions.

ii) Determine the visit ratio to each state of the UBG.

iii) Use the visit ratios to predict the average session length.

iv) Let the demands placed by the requests on the web server be, in seconds

	Home	Add	Buy	Catalog
Web server	0.1	0.2	0.3	0.5

Determine the expected utilization of the web server if new sessions arrive at rate λ sessions/sec. What is the maximum value of λ that the server can sustain before becoming unstable?

b   Give pseudocode for the k-means clustering algorithm. Then, explain how it can help the definition of load tests.

The two parts carry, respectively, 60% and 40% of the marks.

2a A compiler has two optimization flags, A and B. Suppose that a 2^2 design is carried out turning ON and OFF the two factors, simultaneously and one at a time. The measurements collected for the execution time, in seconds, of the compiled program are as follows:

A/B	OFF	ON
OFF	11	9
ON	8	8

i) Propose a response model for this design. Explain the qualitative interpretation of the effects in the proposed model.

ii) Allocate the experimental variation across factors and their interaction.

iii) Suppose that two additional factors C and D, also taking ON and OFF levels, are added to the experiment. The tester is planning to use the following fractional factorial design:

I	A	B	C	AB	AC	D	ABC
1	-1	-1	-1	1	1	1	-1
1	1	-1	-1	-1	-1	1	1
1	-1	1	-1	-1	1	-1	1
1	1	1	-1	1	-1	-1	-1
1	-1	-1	1	1	-1	-1	1
1	1	-1	1	-1	1	-1	-1
1	-1	1	1	-1	-1	1	-1
1	1	1	1	1	1	1	1

where D is treated as a dependent factor. Indicate the name of this design and determine all its confoundings.

iv) What is the resolution of the design in part iii)? Would you recommend using this design?

b Using SPECjbb2015 as an example, present the typical execution phases of an automated benchmarking experiment.

The two parts carry, respectively, 75% and 25% of the marks.

3 a An IT startup is planning to create Alexis, a smart speaker capable of voice interaction. Voice commands given by users to Alexis will be sent to a remote IT backend application. The application will interpret the voice commands and instruct from remote Alexis on how to respond to the user.

The startup manager expects the business to expand rapidly. He is unsure whether the IT backend application should be hosted in the cloud or in the company data centre, which does not use virtualization and can be expanded over time.

i) As the lead performance engineer of Alexis, what case would you present to convince the manager to use a public cloud deployment for the IT backend application rather than the company data centre?

ii) What architecture should the backend application use in order to benefit from horizontal scaling in the cloud?

iii) Give an example of an autoscaling rule based on static thresholds that could help the performance management of the backend application.

b For each of the following statements, state if it is true or false. Give a short justification of your answers.

i) In an IT system, the mix of active requests determines the bottleneck resource.

ii) A CPU bottleneck in a virtual machine can always be mitigated by scaling up the number of cores in the virtual machine.

iii) An IT system always has multiple potential bottlenecks.

iv) Bottleneck switches are beneficial for scalability.

The two parts carry, respectively, 60% and 40% of the marks.