February 26th, 2019

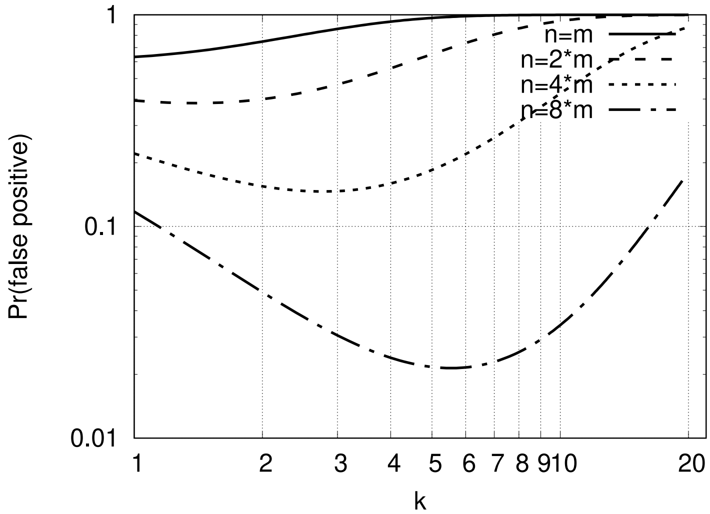
Exam of Switching technologies for data centers (2018/19)

Rules for the exam. It is **forbidden** to use notes, books or calculators. Use only draft paper provided by the professor. When needed, use approximations. The answers must be provided in correct English. Any notation must be defined.
Time available: 70 minutes.

Problem A

Consider a Bloom filter with 800 kbits available for storage and 4 hash functions.

1. Describe which data structure a Bloom filter implements.
2. Describe some possible scenarios in networks where bloom filters are adopted.
3. Write in the pseudocode the operations of “write” and “search” in such bloom filter.
4. Define the probability of false positive.
5. Consider the probability of false positive shown below. After having defined *k*, *n* and *m*, explain the meaning of the graph applied to the specific Bloom filter considered in the question.



Problem B

Consider an input queued switch with Virtual Output Queueing fed by admissible Bernoulli i.i.d. (ABIID) traffic. The adopted scheduler computes the maximum size matching at each timeslot.

1. Explain in details the meaning of “admissible” traffic.
2. Explain in details the meaning “Bernoulli i.i.d.” traffic.
3. Describe an ABIID traffic pattern under which the scheduler achieves 100% throughput.
4. Describe an ABIID traffic pattern under which the scheduler does not achieve 100% throughput and prove such property, by explaining all the required assumptions.

Problem C

Consider the design of a massively large data center based only on Ethernet switches, assuming an oversubscription ratio $3:1$ between server capacity and network capacity. All the adopted switches are equipped with 40 ports running at 10 Gbps. Each server is equipped with one port running at 10 Gbps.

1. For each of the following design problems, draw at high-level the network, compute the required number of switches and the maximum number of supported servers:
   (a) Design the largest possible 2-levels (leaf-and-spine) data center.
   (b) Design the largest possible 2-levels (leaf-and-spine) POD.
   (c) Design the largest possible 3-levels (POD-and-spine) data center.
   (d) Design the largest possible 3-levels (POD-and-spine) POD.
   (e) Design the largest possible 4-levels (POD-and-spine) data center.
2. Based on above results, design a data center connecting 120,000 servers.
3. As summary, fill the following:

Network	Max num. servers	Num. spine ports	Tot. num. switches
2-levels data center			
2-levels POD			
3-levels data center			
3-levels POD			
4-levels data center			
-	$120,000$		

4. Discuss how this design is similar to Google's Jupiter data centers.

Hints for the solution

Problem A

5. k is the number of hash functions. n is the memory size. m is the number of elements already stored in the Bloom filter.
In this specific filter considered in the question, $k = 4$ and $n = 800$ kbit. Thus, the graph shows that

- after inserting $m = 1\,10^5$ elements (i.e., $n = 8m$), the probability of false positive is 0.023;
- after inserting $m = 2\,10^5$ elements (i.e., $n = 4m$), the probability of false positive is 0.16;
- after inserting $m = 4\,10^5$ elements (i.e., $n = 2m$), the probability of false positive is 0.56;
- after inserting $m = 8\,10^5$ elements (i.e., $n = m$), the probability of false positive is 0.93.

Thus, after storing 100 thousands elements in the Bloom filter, the probability of false positive starts to become relatively large (i.e., approximatively larger than 0.1).

Problem C

3.

Network	Max num. servers	Num. spine ports	Tot. num. switches
2-levels data center	$30 \times 40 = 1,200$	-	$40 + 10 = 50$
2-levels POD	$30 \times 20 = 600$	$20 \times 10 = 200$	$20 + 10 = 30$
3-levels data center	$600 \times 40 = 24,000$	-	$40 \times 30 + 200 = 1,400$
3-levels POD	$600 \times 20 = 12,000$	$20 \times 200 = 4,000$	$20 \times 30 + 200 = 800$
4-levels data center	$12,000 \times 40 = 480,000$	-	$40 \times 800 + 4,000 = 36,000$
-	$12,000 \times 10 = 120,000$	-	$10 \times 800 + 1,000 = 9,000$

The last case is obtained by exploiting the 4-levels data center and aggregating 4 spine switches (with 30 unused ports each) into a single one.