Ph.D. Qualify Examination 2020
Theory of Computation

- This examination is closed books.
- Please turn off your cell phones.
- Remember that there are 2 pages of the qualify examination.
- Answer all questions as possible. You may have a partial score if you answer the correct direction.

1. Deterministic Finite Acceptor (DFA) (10 pts)

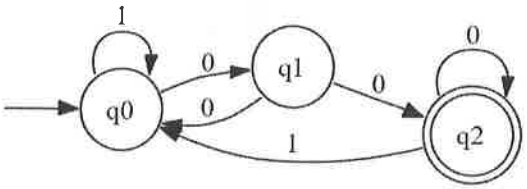
   $L = \{w$: the leftmost symbol differs from the rightmost one for the set of strings$\}$ on $\Sigma = \{0, 1\}$, where $|w| \geq 2$.

2. Nondeterministic Finite Acceptor (NFA) (10 pts)

   Find an NFA with three states that accepts the language

   $L = \{a^n : n \geq 1\} \cup \{b^m a^k : m \geq 0, k \geq 0\}$.

3. Convert the following NFA into an equivalent DFA: (10 pts)



4. Please use the pumping lemma to show that the language is nonregular: (10 pts)

   $L = \{a^n b^l : n/l$ is an integer$\}$.

5. Find a regular grammar that generates the language on $\Sigma = \{a, b\}$ consisting of all strings with no more than two a's. (10 pts)

6. Show that the following grammar is ambiguous. (10 pts)
   $S \rightarrow aSbS|bSaS|\lambda$.

7. Construct an NPDA that accepts the following language on {a,b,c} (use an NPDA with 3 states): (10 pts)

   $L = \{a^n b^m : n \leq m \leq 3n$, where $n \geq 0\}$

8. Prove that all finite languages are regular. (10 pts)

9. Transform the grammar with productions
   $S \to abAB$,
   $A \to baB|\lambda$,
   $B \to BAa|A|\lambda$.
   into Chomsky normal form. (Remove all unit-productions, all useless productions, and all λ-productions from the grammar first) (20 pts)

計算機組織資格考題（Spring 2020）

1. (30%) Refer the following instruction sequence:

Instruction sequence	
lw	$1,40($2)
add	$2,$3,$3
add	$1,$1,$2
sw	$1,20($2)

   a. Find all data dependences in this instruction sequence.
   b. Find all hazards in this instruction sequence for a 5-stage pipeline with and without forwarding.
   c. To reduce clock cycle time, we are considering a split of the MEM stage into two stages. Repeat (b) for this 6-stage pipeline.

2. (20%) Explain the following synchronization primitives: atomic exchange, test-and-set, and fetch-and-increment. Also, explain the following pair of instructions, load linked (LL) and store conditional (SC) and how this pair of instructions can be used to implement atomic exchange and fetch-and-increment.

3. (20%) With dynamic hardware prediction for reducing branch costs, what is the disadvantage of a simple 1-bit branch-prediction buffer for a branch that is almost always taken? Explain why the 2-bit prediction scheme can remedy this disadvantage. Also, explain what correlated predictors are by illustrating an example.

4. (30%) Cache block size (B) can affect both miss rate and miss latency. Assuming a 1-CPI machine with an average of 1.35 references (both instruction and data) per instruction, help find the optimal block size given the following miss rates for various block sizes. (Hint: Average Memory Access Time (AMAT) = (Time for a Hit) + (Miss Rate) x (Miss Latency)).

Block Size	Miss Rate
8	4%
16	3%
32	2%

Block Size	Miss Rate
64	1.5%
128	1%

   a. What is the optimal block size for a miss latency of 20×B cycles?
   b. What is the optimal block size for a miss latency of 24+B cycles?
   c. For constant miss latency, what is the optimal block size?

Qualify Exam – Data Mining, 2020

1. (20 points) Please briefly describe the following terminologies. (1) F1 measure (2) specificity (3) Apriori property (in Apriori Algorithm), (4) False Negative.

2. (20 points) What is "overfitting" and "underfitting" problem in classification modeling? Please also explain how to reduce their effects when you are training models in DNN and decision tree, respectivally.

3. (20 points) Please describe what is semi-supervised learning. If we have only positive and unlabeled data, how do we create a supervised classification model from these data?

4. (20 points) Please apply FP-growth algorithm to find large itemsets in the following transaction data, if mini_support=3.

TID	Items bought
100	{a, c, d, f, g, i, m, p}
200	{a, b, c, f, i, m, o}
300	{b, f, h, j, o}
400	{b, c, k, s, p}
500	{a, c, e, f, l, n}

5. (20 points) A simple labeled data with 4 attributes shown in the right table. Please use **naïve Bayes** method to calculate the class probability of a test instance with "Give Birth"=no, "Can Fly"=no, "Live in Water"=no, and "Have Legs"=yes.

Name	Give Birth	Can Fly	Live in Water	Have Legs	Class
human	yes	no	no	yes	mammals
python	no	no	no	no	non-mammals
salmon	no	no	yes	no	non-mammals
whale	yes	no	yes	no	mammals
frog	no	no	sometimes	yes	non-mammals
komodo	no	no	no	yes	non-mammals
bat	yes	yes	no	yes	mammals
pigeon	no	yes	no	yes	non-mammals
cat	yes	no	no	yes	mammals
leopard shark	yes	no	yes	no	non-mammals
turtle	no	no	sometimes	yes	non-mammals
penguin	no	no	sometimes	yes	non-mammals
porcupine	yes	no	no	yes	mammals
eel	no	no	yes	no	non-mammals
salamander	no	no	sometimes	yes	non-mammals
gila monster	no	no	no	yes	non-mammals
platypus	no	no	no	yes	mammals
owl	no	yes	no	yes	non-mammals
dolphin	yes	no	yes	no	mammals
eagle	no	yes	no	yes	non-mammals

Algorithms 資格考 2020 Spring

1. (30%) Answer each part TRUE or FALSE for the little o notation.
   a) $n=o(2n)$.
   b) $2^n = o(n^2)$.
   c) $2^n = o(3^n)$.
   d) $1 = o(n)$.
   e) $n = o(\log n)$.
   f) $1 = o(\frac{1}{n})$.

2. (25%) Show that the lower bound of sorting.

3. (20%) Give asymptotic tight bounds for $T(n) = 5T(\frac{n}{2}) + \Theta(n^2)$.

4. (25%) Describe a $\Theta(n \lg n)$-time algorithm that, given a set S of n integers and another integer x, determine whether or not there exist two elements in S whose sum is exactly x.

OS 資格考題 (108 學年度第二學期)

1. [15%] Which of the followings are shared among threads in a process? **Explain** your answers briefly.
   (a) program counter
   (b) thread-specific data
   (c) text section
   (d) heap
   (e) stack

2. [15%] Please describe UNIX **system calls**, UNIX **signals**, and their **difference(s)**.

3. [15%] Please describe the 9-bit file access control in UNIX.

4. [15%] What is the benefit of using a **reader-writer** spinlock, compared with a traditional spinlock?

5. [20%] What are the advantages and disadvantages between a **single-level** paging scheme and a **multiple-level** paging scheme?

6. [20%] Consider a demand-paging system with the following **time-measured utilizations**:

CPU	5%
Swap Partition	97%
Other I/O Devices	6%

   Will the following methods improve the CPU utilization? **Briefly explain your answers.**
   (a) Execute more processes
   (b) Enlarge the main memory
   (c) Enlarge the swap partition
   (d) Place the swap partition in a faster disk

CRYPTOGRAPHY (密碼學）資格考

Ph.D Qualify Examination May, 2020

1. Can encryption techniques be used to defend a computer system from the attack of malicious software? How? Why not? (20%)

2. Consider the following protocol: Alice (A) and Bob (B) are users who want to obtain key discussions and mutual authentication from the help of an untrusted sever(S). Kas and Kbs are keys pre-shared between A,S and B,S respectively. Na is a random number produced by A. T is a time stamp (assuming we can have a synchronization clock). Kab is a temporary key generated by S. K will be the key the protocol tries to distribute between A and B. Step1. A sends to S: A, B and Na encrypted in the key Kas. Step2. S sends to B: Na and Kab, both encrypted with the key Kbs. S also sends to A: Na and Kab, both encrypted with the key Kas. Step 3. Bob sends to A: Na, T and K which are all encrypted with the key Kab. Now the question: Is it a valid protocol? Why? Why not? (20%)

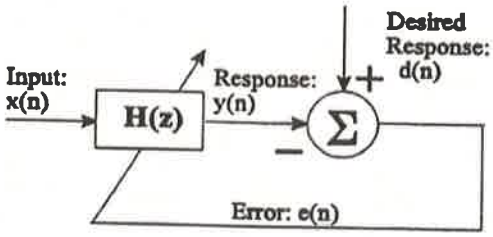
3. Can you design a more efficient one for the problem 2? What assumptions do you need to have? (20%)

4. What is post-quantum cryptography? Give examples please. (20%)

5. How to judge the security level of a crypto system? Give an example to explain. (20%)

Biomedical Signal Processing

1. (25%) For an adaptive filter as shown in Fig. 1, derive the negative gradient of the error function with respect to $b_n(k)$ at time step n. Assume the error function is the sum of squared error between the filter output y(n) and the desired output d(n).



$$y(n) = \sum_{k=0}^{L-1} b_n(k)x(n-k)$$

Fig. 1

2. (25%) Analog-to-digital converter (ADC) converts an analog voltage to an equivalent digital number. (a) What is the resolution of a 5-volt and 12-bit ADC system? (b) Explain the aliasing effect and propose two strategies to deal with aliasing. Give examples.

3. (30%) A group, or ensemble, of time responses averaged together on a point to point basis means ensemble averaging or synchronized averaging. (a) Given two essential requirements to apply ensemble averaging. For the measure y (t) is the signal $x(t)$ with additive random noise η (t) of zero mean. $y(t) = x(t) + \eta(t)$. The signal and noise are statistically independent. (b) Derive $E[y] = u_y = E[x] = u_x$. (c) Derive $E[(y - u_y)^2] = \sigma_y^2 = \sigma_x^2 + \sigma_\eta^2$.

4. (20%) The function of a filter is to retain the components in certain frequency ranges and reject components in other ranges. There are various types of filters such as low-pass, high-pass, band-pass filters, etc. For the noisy ECG signals in Figs. 4-1 and 3-2, which type of filter should be applied to each noisy signal? Explain the frequency characteristics of the signal and noise for each case.

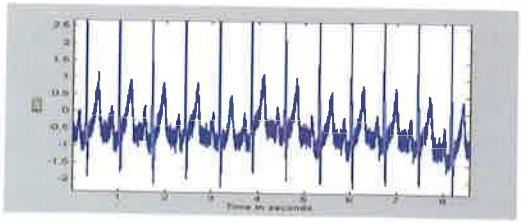


Fig. 4-1

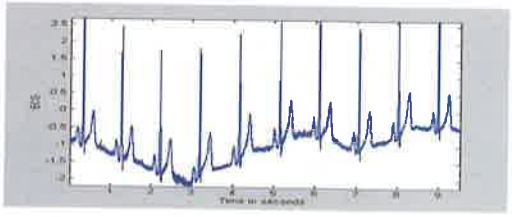


Fig. 4-2

DBMS Qualify Exam
2020 Spring

1. (20%) Assume that we have two relations as follows.

R = R(A, B, C)

S = S(D, E, F)

Give the SQL expression that is equivalent to each of the following queries.

(a) $\Pi_{A,B}(\sigma_{C=D}(R \times S))$

(b) R - S

(c) $\Pi_{A,B}(R) \div \Pi_{D}(S)$

And give the equivalent relational algebra of the following query.

(d) Select A, D From R, S.

2. (15%) A relation, R(A, B, C, D, E, F, G), whose attributes satisfy the functional dependencies:

(BC → A, D, E, F, G), (C → E), (D → F, G), (A→B)

Normalize the above relation to make it satisfy

(a) 2NF

(b) 3NF

(c) BCNF

Note: Don't make unnecessary normalization unless it is required.

3. (45%) Answer the following query in SQL using the given schema:
   S(S#, Sname, Status, City) /* This is a relation for Supplier */
   P(P#, Pname, Color, Weight, City) /* This is a Part relation */
   J(J#, Jname, City) /* This is a Project relation */
   SPJ(S#, P#, J#, Quantity)

   (a) (5%) Get the supplier name and quantity for a part that is supplied by a supplier located at 'Taipei' and its weight is more than 50.
   (b) (10%) Get the name of a supplier who does not supplier any part to a project.
   (c) (10%) Get the total number of projects supplied by supplier S1.
   (d) (10%) Get the supplier names for suppliers supplying all parts that are used in project J1.
   (e) (10%) For each project, get the total quantity and average weight of the parts used in the project.

4. (10%) Explain the following terms.
   (a) Two-phase locking protocol.
   (b) Restrictions to update on views.

5. (10%) Give a formal definition to the relational join operation.

109. 5. 8 VLSI

1. Explain the following terms in detail: (60%)
   (a) critical path (b) setup time
   (c) hard IP (d) functional simulation
   (e) hold time (f) soft IP

2. Describe the difference between full custom and Cell-based design flow. (20%)

3. Suppose you have completed a circuit design with hardware description language. Please describe the advantages and disadvantages of implementing your circuit with (a) ASIC and (b) CPLD/FPGA, respectively. (20%)