

DBMS Qualify Exam 2013 Spring

1. (35%) 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) Get the average weight of the parts that are supplied by each supplier.  
(b) Get the total number of projects supplied by supplier S1.  
(c) Get the supplier names for suppliers supplying all parts that are used in project J1.  
(d) (5%) Answer Question 1(a) in relational algebra.
2. (10%) R(A, B, C, D, E, F) is a relation with functional dependencies:  
(A → B), (B → D), (E → F), (DE → ABCF).  
Normalize the relation to  
(a) 2NF relations  
(b) 3NF relations
3. (15%) Answer the following questions.  
(a) What is the two-phase locking protocol?  
(b) Explain why the two-phase locking protocol can guarantee serializability.  
(c) Explain why the time-stamping protocol can guarantee serializability.
4. (20%) Answer the following questions.  
(a) What is the social network?  
(b) Give two research directions (i.e., research subareas) in social network, and explain why. /\* Note that these subareas don't have to be the ones in the literature. You may give anything you have in mind, as long as you can justify that they are potential research directions. \*/  
(c) What is the cloud computing?  
(d) Give two research directions (i.e., research subareas) in cloud computing, and explain why. /\* Note: Same as above. \*/

More questions on the back

背面有題

5. (20%) Given some relations and the referencing relationships among attributes.

R(r1, r2, r3, r4)            /\* s3 references r1 \*/

S(s1, s2, s3)                /\* s1 references r2 \*/

T(t1, t2, t3, t4)           /\* u1 references s1 \*/

U(u1, u2, u3, u4)         /\* u2 references t1 \*/

Give an educational guess what the corresponding ER schema of this database would be. Draw this ER schema diagram, indicate the cardinality ratio (such as 1:1, 1:N, M:N, etc.), and explain why.

10/2012 博士班資格考: 機率與統計 Show All Details.

1. (20%) Let  $X$  and  $Y$  be two RVs.  $X$  is a Poisson distribution with parameter  $\lambda$ . If  $y = x^2$ , find the *p.d.f.* of  $Y$ .

2. (20%) Let  $X$  and  $Y$  be two random variables with uniform distribution in the interval  $[0,2]$ . Let  $Z=2X+Y$ . Find the probability distribution of  $Z$ .

3. (20%) The possible number of defective items produced from a factor in one production run is modeled by a binomial distribution. (20%)

(a) Write down the density function of a binomial distribution.

(b) We know *a priori* that the probability is 0.7 for the case that 1 out of 5 is defective and the probability is 0.3 for the case that 2 out of 5 are defective. For the current production run, we find that 1 out of 3 are defective from random samples. Find the Bayes estimate of the proportion of the defectives for this production run.

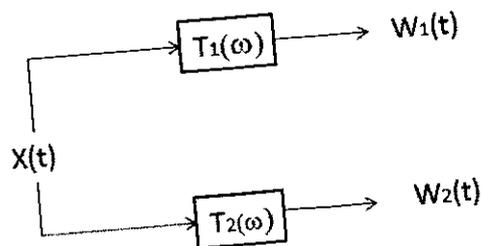
4. (20%)

(a) State the differences among independent, uncorrelated, and orthogonal

(b) What is Strict Sense Stationary (SSS) and what is Wide Sense Stationary?

5. (20%) A random process is called  $\sigma$ -bandlimited if it has finite power and its spectrum,  $S(\omega)$ , vanishes for the frequency larger than  $\sigma$ . That is,  $S(\omega) = 0$ , for  $|\omega| > \sigma$ .

Given the following system. Show that if  $T_1(\omega) = T_2(\omega)$  for  $|\omega| \leq \sigma$ , then  $w_1(t) = w_2(t)$ .



**2013 April NCKU CSIE PH.D. Qualification Examination**  
**Computer Architecture**

1. Some operations on two operands (e.g., subtraction) are not commutative. What are the advantages and disadvantages of the stack, accumulator, and load-store architectures when executing noncommutative operations? (15 points)
2. The classical approach to improving cache behavior is to reduce miss rates. Please summarize the techniques that can reduce miss rates. (15 points)
3. Explain why the cache miss penalties increase as the processor becomes much more faster than DRAMs? Briefly describe five optimizations to reduce cache miss penalty. (20 points)
4. Describe two major instruction set characteristics that can further divide general purpose register (GPR) instruction set architecture into three classes, based on whether the instruction operands are used explicitly or implicitly. And show the advantages and disadvantages of these three further divided classes. (20 points)
5. Describe the advantages and disadvantages of dynamic scheduling? What are the key concept of Tomasulo's approach? (15 points)
6. Describe what are the RAW, WAW, and WAR hazards. (15 points)

Ph.D. Qualify Examination 2013  
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 Acceptors (DFAs) (10 pts)  
Draw Deterministic Finite Automata to accept the set of all strings over  $\{a, b\}$  that do not contain the substring "aab".
2. Nondeterministic Finite Acceptors (NFAs) (20pts. 10 pts each)  
Draw Non-deterministic Finite Automata to accept the following sets of strings over  $\{a, b\}$ :
  - (a) All strings end in "aaba".
  - (b) All strings where every odd position is "a".
3. Prove that the language is nonregular: (15 pts)  
 $L = \{a^n b a^n : n \geq 0\}$ .
4. Suppose  $r_1$  and  $r_2$  are regular expressions, and  $L(r_1)$  and  $L(r_2)$  are regular languages, correspondingly. Prove that  $L(r_1 + r_2)$  is also a regular language (15 pts).  
hint: Note that  $L(r_1 + r_2) = L(r_1) \cup L(r_2)$ . You can consider the solution by using NFA.
5. Show that the following grammar is ambiguous. (10 pts)  
 $S \rightarrow AB|aaB$   
 $A \rightarrow a|Aa$   
 $B \rightarrow b$
6. Construct a nondeterministic pushdown automata that accept the following language over  $\Sigma = \{a, b\}$ : (10 pts)  
 $L = \{w \in \{a, b\}^* : n_a(w) = n_b(w)\}$ , where  $n_a(w)$  is the count of "a" in  $w$  and  $n_b(w)$  is the count of "b" in  $w$ .

7. Fill the following languages into the language hierarchy (If  $L_i$  is a regular language and also a context-free language, please fill  $L_i$  in the set of regular languages): (20 pts)

$$L_1 = \{a^n b^n : n \geq 0\},$$

$$L_2 = L(a^* b^*),$$

$$L_3 = \{a^n b^n c^n : n \geq 0\},$$

$$L_4 = \{a^n w w^R a^n : n \geq 0, w \in \{a, b\}^*\},$$

$$L_5 = \{ab, ad, a\},$$

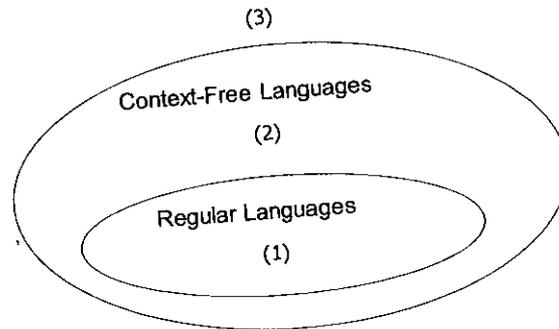
$$L_6 = \{ww : w \in \{a, b\}^*\},$$

$$L_7 = \{a^{n!} : n \geq 0\},$$

$$L_8 = \{a^n b^j a^j b^n : n \geq 0, j \geq 0\},$$

$$L_9 = \{a^n b^m c^{n+m} : n \geq 0, m \geq 0\},$$

$$L_{10} = \{a^3 b^n c^n : n \geq 0\}.$$



## 2013 April NCKU CSIE PH.D. Qualification Examination

### 高效能網路技術

1. For a given prefix set  $\{0^*, 000^*, 010^*, 01011^*, 011^*, 11^*, 111^*\}$ ,
  - (a) Please draw the binary tries. (10%)
  - (b) Please list the most specific prefixes. (5%)
  - (c) Please find the longest matching prefix for address 01011010(5%)
  
2. Binary Prefix Search (BPS) is an efficient IP address lookup scheme that provides us a comparison method to sort prefixes. According to BPS and the same prefix set in question 1,
  - (a) Please sort the prefixes. (10%)
  - (b) Please give an example to illustrate why it might cause a failed lookup when we directly perform a binary search on these sorted prefixes. (5%)
  - (c) How to perform IP lookup by using BPS? (10%)
  
3.
  - (a) Please list the pro's and con's of Ternary Content Addressable Memories (TCAMs). (10%)
  - (b) Please convert a 4-bit range [1, 14] to the format that can be stored in TCAMs. (10%)
  
4. In packet classification problem, what are the differences between the header cache and the rule cache? (15%)
  
5. Let R be a set of five 5-bit ranges. Where  $R = \{[3, 15], [7, 7], [20, 25], [8, 19], [25, 31]\}$ .
  - (a) Please list all the elementary intervals induced by R. (10%)
  - (b) Please draw a balanced binary segment tree that constructed according to R. (10%)

Algorithms 資格考 2013 Spring

1. (30%) Solve  $T(n) = 9T(\frac{n}{3}) + n$  using  $\Theta$ .
2. (20%) (a)(10%) Define  $O(g(n))$ . (b)(10%) Define  $\Theta(g(n))$ .
3. (20%) Present a linear-time algorithm to compute strongly connected components of a directed graph.
4. (30%) Show the lower bound of any comparison sort algorithm.

# Discrete-Time Signal Processing 資格考

April 2013

1. (20%) Determine if the systems described by the following input-output equations are (1) linear, (2) stable, and (3) causal.

- (a)  $y[n] = 3x[n] + 5$   
 (b)  $y[n] = \log(x[n-5])$   
**Justify your answer.**

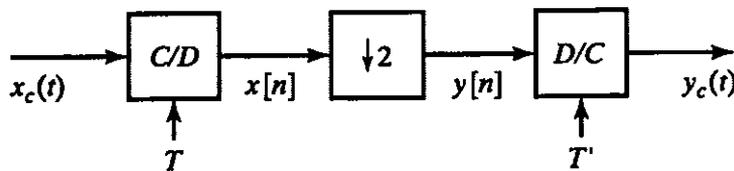
2. (20%) In the following figure,  $x[n] = x_c(nT)$  and  $y[n] = x[2n]$

- (a) Assume that  $x_c(t)$  has a Fourier transform such that  $X_c(j\Omega) = 0, |\Omega| > 2\pi(100)$ .

What value of  $T$  is required so that

$$X(e^{j\omega}) = 0, \quad \frac{\pi}{2} < \omega \leq \pi ?$$

- (b) How should  $T'$  be chosen so that  $y_c(t) = x_c(t)$ ?



3. (20%) Consider a right-sided sequence  $x[n]$  with z-transform

$$X(z) = \frac{2z^2 - z}{2z^2 + \frac{3}{2}z + \frac{1}{4}}$$

Determine the inverse z-transform using each of the following methods

4. (20%) Consider a stable linear time-invariant system with input  $x[n]$  and output  $y[n]$ . The input and output satisfy the difference equation.

$$y[n-1] - \frac{5}{2}y[n] + y[n+1] = x[n]$$

- (a) Plot the poles and zeros in the z-plane.  
 (b) Find the impulse response  $h[n]$ .

5. (20%). Suppose that we wish to design a highpass filter satisfying the following specification:

$$-0.04 < |H(e^{j\omega})| < 0.04, \quad 0 \leq \omega \leq 0.2\pi,$$

$$0.995 < |H(e^{j\omega})| < 1.005, \quad 0.3\pi \leq \omega \leq \pi.$$

The filter will be designed using the bilinear transformation and  $T=2$  ms with a prototype continuous-time filter. State the specification that should be used to design the prototype continuous-time filter to ensure that the specification for the discrete-time filter are met. The following equations are for your reference :

$$s = \frac{2}{T} \left( \frac{1-z^{-1}}{1+z^{-1}} \right), \quad \Omega = \frac{2}{T} \tan(\omega/2), \quad \omega = 2 \arctan(\Omega T / 2)$$

## 作業系統 資格考題 (101 年度第 2 學期)

1. (20%) Briefly describe the content/fields in an inode.
2. (20%) Please write the pseudo code of a multi-threaded web server. You may need to use the following APIs.
  - Socket API: `socket()`, `connect()`, `accept()`...
  - Pthread API: `pthread_create()`, `pthread_join()`...
  - File system API: `read()`, `write()`...
3. (20%) Explain the difference between internal fragmentation and external fragmentation. Which one occurs in paging system? Which one occurs in systems using pure segmentation?
4. (20%) Explain the four necessary conditions for a deadlock to occur, and suggest a method to break one of the conditions so as to prevent deadlocks.
5. (20%) A computer with a 32-bit address uses a two-level page table. Virtual addresses are split into a 10-bit top-level page table field, an 10-bit second-level page table field, and an offset.
  - (a) What's the page size?
  - (b) If a program has its code and data together fitted in the lowest 4KB and stack fitted in the highest 4KB of its virtual address space, how many pages are needed for storing the page table for this program?

# 密碼學博士資格考

(Close book, 2013/4/12)

(1). (50%) Alice wants to send an important data file  $F$  of 8000 M bytes to a remote friend Bob via a public network. Please describe ways to provide

- (a) privacy to  $F$ ;
- (b) privacy and integrity to  $F$ ;
- (c) privacy in realtime to  $F$ ;
- (d) privacy and integrity in realtime to  $F$ ;
- (e) privacy using quantum mechanics.

(2) (50%) What is Quantum Foulrier Transform?

Describe its usefulness in Quantum Computing and Quantum Cryptography.

## Qualify Exam - Data Mining

Notice: Close book. 總分: 100 分

(共 1 頁)

1. (16%) Answer the following questions:
  - A. (6%) Given a dataset  $D$  of  $m$  records, with attributes  $A_1, A_2, \dots, A_n$  in each record, where attribute  $A_i$  ( $1 \leq i \leq n$ ) could be numerical or categorical type. Give an effective method for dealing with the missing values in  $D$  without deleting the records with missing values.
  - B. (5%) Give the formal definition of "Association Rule".
  - C. (5%) Give the equation of *Lift* and explain how it can be used to measure the interestingness of association rules.
  
2. (22%) Answer the following questions:
  - A. (12%) Give the algorithm of *Apriori* for mining frequent itemsets and explain how *Apriori* performs pruning on candidate itemsets.
  - B. (10%) Design an efficient method for mining frequent itemsets without generating candidates.
  
3. (30%) Answer the following questions on clustering methods:
  - A. (12%) As comparisons, give advantages and drawbacks for each of the following type of clustering methods: "Partition-based clustering", "Hierarchical clustering" and "Density-based clustering".
  - B. (12%) *BIRCH* is an efficient clustering method. Which type of clustering (Partition, Hierarchical and Density-based) does *BIRCH* belongs to? Give the algorithm of *BIRCH* and explain its advantages.
  - C. (6%) Explain how to calculate the similarity between two data items for clustering if there exist both of numerical and categorical attributes in each data item.
  
4. (32%) Answer the following questions about classification modeling:
  - A. (12%) C4.5 is a popular method for classification modeling. Give the algorithm of C4.5 and explain what kind of splitting measure is used in C4.5.
  - B. (10%) Describe how "Ensemble" method works for building a classifier and explain why it can reach higher accuracy normally.
  - C. (10%) Given a dataset  $D$  with highly imbalanced classes distribution and low classification accuracy. Give the best pre-processing approach you think for dealing with  $D$  such that the classification accuracy can be improved most.

# Computer Networks – Qualified Exam

(2013/Spring)

系別 \_\_\_\_\_ 年級 \_\_\_\_\_ 學號 \_\_\_\_\_ 姓名 \_\_\_\_\_

1. (10%)

Let a link's distance be 2,500 km, propagation speed be  $2.5 \cdot 10^8$  m/s, and transmission rate be 2 Mbps. How long does a packet of length 1,000 bytes take to propagate over this link? Let's derive the general form: how long does it take a packet of length  $L$  to propagate over a link of distance  $d$ , propagation speed  $s$ , and transmission rate  $R$  bps? Does this delay depend on the packet length? Does this delay depend on the transmission rate?

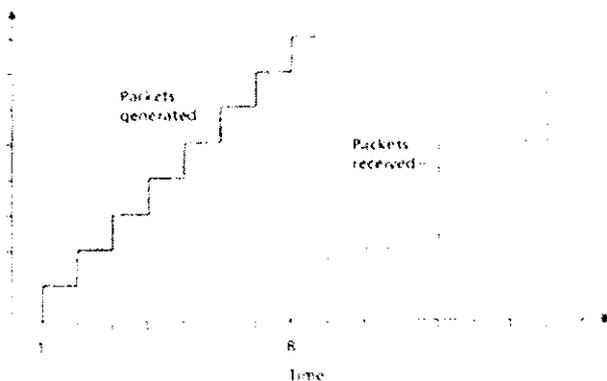
2. (10%)

i) HTTP, FTP, SMTP, and POP3 run on top of TCP. Why aren't they on the top of UDP?

ii) Let Alice provide chunks to Bob throughout a 30-second interval using BitTorrent. Will Bob necessarily return the favor and provide chunks to Alice in this same interval? Why or why not?

3. (10%)

Consider the figure below. A sender begins sending packetized audio periodically at  $t = 1$ . The first packet arrives at the receiver at  $t = 8$ .



i) What are the delays (from sender to receiver, ignoring any playout delays) of packets 2 through 8? Note that each vertical and horizontal line segment in the figure has a length of 1, 2, or 3 time units.

ii) If audio playout begins as soon as the first packet arrives at the receiver at  $t = 8$ , which of the first eight packets sent will not arrive in time for playout?

iii) If audio playout begins at  $t = 9$ , which of the first eight packets sent will not arrive in time for playout?

iv) What is the minimum playout delay at the receiver that results in all of the first eight packets arriving time for their playout?

4. (10%)

Consider a datagram network using 8-bit host addresses. Suppose a router uses longest prefix matching and has the following forwarding table:

Prefix	Interface
00	0
01	1
10	2
11	3

For each of the four interfaces, give the associated range of destination host addresses and the number of addresses in the range.

5. (10%)

Suppose three active nodes—nodes A, B, and C—are competing for access to a channel using slotted ALOHA. Assume each node has an infinite number of packets to send. Each node attempts to transmit in each slot with probability  $p$ . The first slot is numbered slot 1, the second slot is numbered slot 2, and so on.

- What is the probability that node A succeeds for the first time in slot 4?
- What is the probability that some node (either A, B or C) succeeds in slot 2?
- What is the probability that the first success occurs in slot 4?
- What is the efficiency of this three-node system?

6. (10%)

i) In classless addressing, can two blocks have the same prefix length? Explain.

ii) In classless addressing, we know the first address and one of the addresses in the block (not necessarily the last address). Can we find the prefix length? Explain.

7. (10%)

Using 5-bit sequence numbers, what is the maximum size of the send and receive windows for each of the following protocols?

- Stop-and-Wait
- Go-Back- $N$
- Selective-Repeat

8. (10%)

i) In cases where reliability is not of primary importance, UDP would make a good transport protocol. Give examples of specific cases.

ii) Are both UDP and IP unreliable to the same degree? Why or why not?

9. (10%)

i) What is the maximum number of routers that can be recorded if the timestamp option has a flag value of 1? Why?

ii) The value of HLEN in an IP datagram is 7. How many option bytes are present?

10. (10%)

A client uses UDP to send data to a server. The data is 16 bytes. Calculate the efficiency of this transmission at the UDP level (ratio of useful bytes to total bytes).