# 密碼學博士資格考

# Cryptography Ph.D Qualified Exam.

(Close book, 2014/4/17)

Describe the evolution of information security (including cryptography) according to the improvement of science or technology.

**5/2015 博士班資格考: 機率與統計 Show All Details.**

1. (30%)

   Given a distribution with probability density function $f(x)=\lambda e^{-\lambda x}$, for $\lambda>0$.

   (a) Find the moment generation function of $x$.

   (b) Use the moment generation function to find the mean and variance of the distribution.

   (c) Let $y=x^2$, find the moment generation function of $y$.

   (d) Use the above result to find the probability density function of $y$.

2. (25%)

   Let $X_1, X_2, \cdots, X_N$ denote the independent random variables taken from a discrete probability distribution, $f(X,\theta)$, where $\theta$ is a single parameter of the distribution.

   (a) Define the likelihood function.

   (b) Explain the function of a maximum likelihood estimator.

   (c) For a Normal distribution $N(\mu,\sigma)$, find the maximum likelihood estimators for $\mu$ *and* $\sigma$.

3. (15%)

   Let **y** is estimated by **ax,** where **y** and **x are RVs.** Show that the MS error $e=E\{(y-ax)^2\}$ is minimized when a is such that $E\{(y-ax)x\}=0$

4. (15%)

   Let X be a random variable and Y=g(X). Prove that

   (a) E[Y] = E[g(x)].

   (b) $f_Y(y)=\frac{f_X(x_1)}{|g'(x_1)|}+\frac{f_X(x_2)}{|g'(x_2)|}+\cdots+\frac{f_X(x_n)}{|g'(x_n)|}+\cdots$, where $y=g(x_1)=g(x_2)=\cdots$

5. (15%)

   A random sample of 40 students obtained a mean of 75 and a variance of 15 on a college placement test in math. Assume the scores to be normally distributed, construct the 98% confidence interval for $\sigma^2$.

# 圖論資格考 2015/4

1. (10%) For a set $S \subseteq N$ of size $n$, determine the number of trees with vertex set $S$.
2. (10%) Show that every connected graph contains a spanning tree.
3. (20%) Show that a graph is bipartite iff it has no odd cycle.
4. (20%) Determine the Wiener index of an $n$-vertex path.
5. (10%) Show that the center of a tree is a vertex or an edge.
6. (10%) Show that if $k>0$, then a k-regular bipartite graph has the same number of vertices in each partite set.
7. (20%) Show that every component of the symmetric difference of two matchings is a path or an even cycle.

# *Digital Image Processing*
# *Ph.D. Qualification Examination,*
## *Department of CSIE*

*April 2015*

1. (a) What is the 1-D convolution theorem of a continuous variable? (4%) please prove it. (6%) (b) What is aliasing? (4%) Will the shrinking and zooming operations cause aliasing? (Shrinking? zooming? or both?) Why? (6%)

2. You are given a medical image and asked to filter the image in frequency domain. Please give the steps that you use to perform image filtering and explain the function of each step. (12%) Please also explain how you can put the origin of frequency image at the image center and prove it. (8%)

3. A certain color transform in RGB color space is $s_i = kr_i$, $i$=1,2,3, $k$ is the scaling parameter. (a) Please **show** its corresponding transform in CMY space is $s_i = kr_i + (1 - k)$ $i$=1,2,3. (10%) (b) Please also give the formula of the above transform in HSI space, and describe the advantages and disadvantages for using transform in HSI space. (10%)

4. The Laplacian of Gaussian filter is given as

$$\nabla^2 G(x,y) = \left[\frac{x^2+y^2-2\sigma^2}{\sigma^4}\right] e^{-\frac{x^2+y^2}{2\sigma^2}}$$

(a) Please prove that the average value of this filter is zero. (Please note

$$\sigma^2 = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} z^2 e^{-\frac{z^2}{2\sigma^2}} dz \quad \text{and} \quad \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-\frac{z^2}{2\sigma^2}} dz = 1$$

, 10%)

(b) Please also show that the average value of an image convolved with this filter is also zero. (Hint: Please using the convolution theorem and DC value in Fourier transform in the proof.) (10%)

5. What is the watershed segmentation algorithm? (6%) What are the common drawbacks of this method? (6%) How will you handle these problems? (7%)

# OS 資格考題 (103 學年度第二學期)

1. (20%) Assume that a demand-paging system has **4 page frames**, and the size of a **page frame is 10 bytes**. The memory is byte-addressable and the memory reference string is as follows:

   Addresses: 70, 09, 18, 24, 01, 32, 02, 40, 22, 30, 02, 30, 23, 18

   Please answer the following questions, and **explain** the answers.
   a. What is the page fault number under the LRU page replacement strategy?
   b. What is the minimal page fault number?

2. (20%) Please describe the advantage(s) and disadvantage(s) of setting a very small time quantum in RR scheduling.

3. (20%) Consider a demand-paging system with the following **time-measured utilizations**:

| | |
|---|---|
| CPU | 5% |
| Swap Partition | 98% |
| Other I/O Devices | 7% |

   Will the following methods improve the CPU utilization? **Briefly explain your answers.**
   a. (5%) Execute more processes
   b. (5%) Enlarge the main memory
   c. (5%) Enlarge the swap partition
   d. (5%) Place the swap partition in a faster disk

4. (15%) What's thread-specific data? What are the difference(s) between thread-specific data, local variables, and global variables?

5. (15%) What are the difference(s) between deadlock prevention and deadlock avoidance? Which one you prefer, and why?

6. (10%) Please describe the benefit(s) of using a reader-writer semaphore, compared to the use of a traditional semaphore.

# Digital Signal Processing 資格考

April 2015

1. **(20%)** Consider an LTI system with frequency response

$$H(e^{j\omega}) = \frac{1-e^{-j2\omega}}{1+\frac{1}{2}e^{-j4\omega}}, \quad -\pi < \omega \le \pi$$

Determine the output $y[n]$ for all $n$ if the input $x[n]$ for all $n$ is

$$x[n] = \sin(\frac{\pi n}{4}).$$

2. **(20%)** A causal and stable LTI system $S$ has its input $x[n]$ and output $y[n]$ related by the linear constant-coefficient difference equation

$$y[n] + \sum_{k=1}^{10} \alpha_k y[n-k] = x[n] + \beta x[n-1],$$

Let the impulse response of $S$ be the sequence $h[n]$

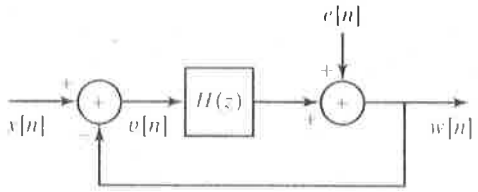
(a) Show that $\alpha_1$ can be determined from the knowledge of $h[0]$ and $h[1]$.

(b) If $h[n] = (0.9)^n \cos(\pi n/4)$ for $0 \le n \le 10$, sketch the pole-zero plot for the system function of $S$, and indicate the region of convergence.

3. **(20%)** In the following figure, H(z) is the system function of a causal LTI system. As shown in the figure, $W(z)$ can be expressed in the form

$$W(z) = H_1(z)X(z) + H_2(z)E(z)$$

For the case $H(z) = \frac{z^{-1}}{(1-z^{-1})}$, determine $H_1(z)$ and $H_2(z)$.




4.**(20%)** A discrete-time causal LTI system has the system function

$$H(z) = \frac{(1+0.2z^{-1})(1-9z^{-2})}{(1+0.81z^{-2})}.$$

(a) Is the system stable?

(b) Determine expressions for a minimum-phase system $H_I(z)$ and an all-pass system $H_{ap}(z)$ such that

$$H(z) = H_I(z)\, H_{ap}(z)$$

5.**(20%)** Let $X(e^{j\omega})$ denote the Fourier transform of the sequence $x[n] = (\frac{1}{2})^n u[n]$. Let $y[n]$ denote a finite-duration sequence of length 10; i.e., $y[n] = 0$, $n < 0$, and $y[n] = 0$, $n \ge 10$. The 10-point DFT of $y[n]$, denoted by $Y[k]$, corresponds to 10 equally spaced samples of $X(e^{j\omega})$ ; i.e., $Y[k] = X(e^{j2\pi k/10})$. Determine $y[n]$.

# Algorithms 資格考 April 2015

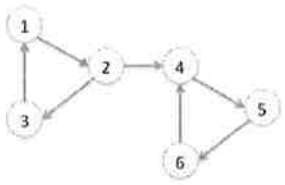
1. (10%) Give formal definitions of $\Theta(g(n)), O(g(n))$, and $\Omega(g(n))$.
2. (10%) Give asymptotic upper and lower bounds for $T(n)$ in the following recurrence: $T(n) = T(\frac{n}{3}) + T(\frac{2n}{3}) + O(n)$ (please make your bound as tight as possible).
3. (10%) Describe a $\Theta(n \lg n)$-time algorithm that, given a set of $n$ integers and another integer $x$, determine whether or not there exist two elements in $S$ whose sum is exactly $x$.
4. (20%) Show that any comparison sort algorithm requires $\Omega(n \lg n)$ comparisons in the worst case.
5. (10%) Present the radix sort algorithm and analyze the complexity.
6. (10%) Present the quick sort algorithm and analyze the complexity.
7. (20%) Show how to sort n integers in the range 0 to $n^3 - 1$ in $O(n)$ time.
8. (10%) (a) (5%) Determine which one of the 0-1 knapsack problem and the fractional knapsack problem cannot be solved using the greedy strategy? (b) (5%) Give an example to explain that.

# Information Retrieval

*PhD Qualification Examination (2105 Spring)*

1. (30 points) Please describe the following terminologies. (1) Ad-hoc search (2) LOOCV (3) probabilistic latent semantic analysis (4) Bayesian network (5) Lemmatization (6) POS tagging (7) Boolean Model (8) Wrapper (9) Break-even point (10) 1-NN

2. (20 points) (a) From this graph in the right figure, please analysis and explain what is the effects for **hub / authority** value of Node 4 if we add a new edge 2→6 (this is also called a shortcut link). (b) Please describe applications / problems in which the mutually reinforcement link algorithms, e.g., HITS, PageRank, are better and more sensitive than general local link algorithms, e.g., in-degree counting.




3. (25 pts) For the example in the following table, the multinomial parameters we need to classify the test document are the priors P(*c*) = 3/4, P(not *c*) = 1/4. Please find the value of P("Chinese"|not *c*), P("Japan"|not *c*), P(*c*|$d_5$) and P(not *c*|$d_5$). [Hint: P("Chinese"|*c*)=(5+1)/(8+6)=3/7]

| | *docId* | *Words in document* | *in c = China?* |
|---|---|---|---|
| Training set | 1 | Chinese, Beijing, Chinese | *Yes* |
| | 2 | Chinese, Chinese, Shanghai | *Yes* |
| | 3 | Chinese, Macao | *Yes* |
| | 4 | Tokyo, Japan, Chinese | *No* |
| Test set | 5 | Chinese, Chinese, Chinese, Tokyo, Japan | ? |

4. (25 pts) You have an experiment dataset, which contains 10 positive cases (P), 30 negative cases (N), and 60 un-labeled cases (U). You have implemented two retrieval systems (M1 and M2) and you want to evaluate which one is better. The following table shows their top-30 retrieval results. Please answer:
   (1). R-Precision and MAP of M2.
   (2). Use MRR of M1 and M2 to compare their performance. Is MRR good for judging your systems?
   (3). DCG of M1, where relevance score (i-th P cases)=11-i, N, U cases score 0.
   (4). Which system is better? How do you judge this? How to improve it?
   (5). What are the effects of these 60 un-labeled cases? Will they make your judgment different?

| Top-30 results | #P (ranks) | #N | #U |
|---|---|---|---|
| M1 | 5 (1, 4, 5, 11, 20) | 15 | 10 |
| M2 | 3 (2, 10, 20) | 7 | 20 |

# DBMS

*PhD Qualification Examination (2105 Spring)*

1. (40 pts) Please describe the following terminologies / concepts. (1) 4NF (2) DBA (3) Views (4) Boyce-Codd Normal Form (BCNF) (5) $B^+$-Tree (6) Cascading Rollback (7) NoSQL (8) Two-Phase Locking Protocol.

2. (30 pots) Suppose we decompose the scheme R = (A, B, C, D, E) into (A, B, C), (A, D, E). (a) Show that this is a lossless-join decomposition if the following set F of functional dependencies hold. (b) Give a lossless-join decomposition of the scheme R into BCNF. (c) Give a lossless-join, dependency-preserving decomposition of scheme R into 3NF.

   A→BC

   CD→E

   B→D

   E→A

3. (30 pts) 一家醫院欲設計電子病歷系統，但對應窗口並無相關資訊技術，僅開出如下規格：

   醫生資料: 科別, 年資

   門診資料: 診別, 時間, 主治醫生,

   病患: 姓名, 個人資料

   就診記錄: ?

   (a) 請根據想像設計一更完整的簡易醫院就診記錄查詢系統(需能夠回答(b)的查詢，並繪出相關 E-R diagram.

   (b) 請根據你設計的資料庫寫出以下相對應的 SQL 查詢字串 (1) 在日期 D 來看內科且有拿藥的六十歲女性 (2) 看過兩種以上科別的病患姓名與對應醫生年資.

**Qualifications Test of Context-aware System Design (2015, open book)**

請基於context awareness原理，提出具創新服務價值的cloud service for health care應用情境，及回答下列問題：

(a)提出實現此情境感知雲端服務的系統架構，並說明其滿足 ubiquitous context-aware service 的特性。另請從網路匯流、感知網路、人機互動、服務創新性、安全信賴、建置及維運成本等面向分析前述構想的關鍵成功因素。(25 分)

(b)請提出實現前述構想之 context model。(20 分)

(c)請依據 TEA context perception architecture，設計前述構想之 context acquisition mechanism。(15 分)

(d)請說明如何採用 context broker paradigm 實現前述構想。若不適合採用此paradigm，亦請說明其原因。(15 分)

(e)請說明如何採用 context widget 原理實現前述構想，並示範 interpreters, aggregators 等 context toolkit 元件在你的設計中可以發揮的功能。(15 分)

(f)請說明前述構想如何採用 Sentient Object Model 來實現 context management(應包括多個 Sentient Objects 互動的情境)，並說明其如何運作。若不適合採用此paradigm，亦請說明其原因。(10 分)

NCKU CSIE Ph.D. Qualification Examination May 2015

Computer Architecture

1. Explain the following terms (20%, 4 points each)
   a. Amdahl law
   b. Delayed branch
   c. Precise exception
   d. Simultaneous multithreading
   e. Cache coherence
2. Answer the following questions (20%, 5 points each)
   a. What is the influence of cache miss rate (increased or reduced) if larger block size is used? Explain why.
   b. What is the influence of cache miss rate (increased or reduced) if higher associativity is used? Explain why.
   c. What is the influence of cache miss penalty (increased or reduced) if multilevel cache is used? Explain why.
   d. What is the influence of hit time (increased or reduced) if trace cache is used? Explain why.
3. (40%) Answer the following questions.
   a. [10%] Explain what RAW, WAW and WAR hazards are
   b. [10 %] Refer the following instruction sequence. Find all data hazards in this instruction sequence for an 5-stage pipeline if data forwarding is **NOT** used. Note that the instructions are executed in order.

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

   c. [10%] Repeat (b) if data forwarding **IS** used.
   d. [10%] To reduce the clock cycle time, we are considering a split of the MEM stage into two stages. Find all data hazards in this instruction sequence for this architecture if data forwarding is NOT used.
4. (20%) With dynamic hardware 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 is by illustrating an example.

# 104. 4. 17 VLSI

**1. Explain the following terms in detail: (60%)**

**(a) critical path** **(b) setup time**
**(c) power dissipation** **(d) clock skew**
**(e) RTL** **(f) hard IP**

**2. Describe the difference between full custom and Cell-based design flow. (15%)**

**3. What is the difference between "a front-end designer" and "a back-end designer" ? (10%)**

**4. Explain the difference of the following two graphic symbols (5%) and draw their timing outputs Q and F (gate delay is ignored) for the input sequence (10%).**