UCLA
Computer Science Department
Winter 2011

Student Name and ID: ________________________________

CS144 Final: Closed Book, 2 hours

(IMPORTANT PLEASE READ **):**

- There are 4 problems on the exam to be completed in 2 hours. *You should look through the entire exam before getting started, in order to plan your strategy.*
- You may use two sheets of double-sided notes during exam. You are also allowed to use a calculator. Attach extra pages as needed. Write your name and ID on the extra pages.
- *Simplicity and clarity of your solutions will count.* You may get as few as 0 points for a problem if your solution is far more complicated than necessary, or if we cannot understand your solution.
- If you need to make any assumption to solve a question, please *write down your assumptions.* You may also want to write down how you arrived at your answer step by step to get partial credit.

Problem	**Score**	
1	60	
2	20	
3	10	
4	10	
Total	100	

Problem 1: 60 points

This problem consists of 15 YES/NO questions. For each statement below, circle CORRECT or INCORRECT. Be careful with your answers, because we will give 4 points for every correct answer and **deduct 2 points for every incorrect answer**.

1. At the beginning of every SSL communication, the server presents to the browser a certificate signed by a certificate authority (CA). In order to check the validity of this certificate, the browser has to communicate with the CA over the Internet.

   ANSWER:
   `INCORRECT. Every browser already has the public keys for trusted CAs, so it can use the public keys to verify the authenticity of the certificate.`

2. We assign formatting properties to Web pages via the Cascading Style Sheet (CSS). The CSS for the page P assigns two conflicting properties to an HTML element e on P through its class and its ID. In this case, the property assigned through its class will be applied to the element e.

   ANSWER:
   `No, the ID overrides class.  The more specific property wins.`

3. The RSA cipher is secure as long as *either* the RSA problem *or* the large-number factorization problem remains to be computationally too expensive a problem.

   ANSWER:
   `INCORRECT, both RSA problem *AND* the large-number factorization problem must remain computationally expensive.`

4. We need to write a Web page that allows users to update their profile on our site through an HTML form. As the method attribute of the form, we are generally allowed to use either GET or POST.

   ANSWER:
   `INCORRECT, when a request may leave an important side effect on the server, we should use POST.`

5. According to the HTTP standard, both the response to a GET request and the response to a POST request may be cached by a Web browser by default.

ANSWER:
INCORRECT, the response from POST should not be cached by the browser. By default, the browser has to issue every POST request from the user unless the server explicitly states that the response can be cached. This is because a POST request may leave an important side effect at the server, so the browser cannot be sure whether ‘‘skipping’’ a POST request will be safe just because the same request was issued earlier.

6. Consider a text whose size is 1KB when the text is encoded in ISO-8859-1 (Latin-1). If the same text is encoded in UTF-16, its size would always be 2KB, ignoring the Unicode byte order mark.

   ANSWER:
   CORRECT, UTF-16 uses 2 bytes for storing any Latin-1 character.

7. For any text that can be represented by ISO-8859-1 character sets, choosing UTF-8 to encode it is always more efficient than choosing the UTF-16 in terms of the size of the resulting document.

   ANSWER:
   CORRECT, UTF-16 would use 2 bytes to represent a Latin-1 character, but the UTF-8 may use 1 or 2 bytes.

8. A browser is able to establish a channel with a server that no one else can eavesdrop even if the server's certificate is not signed by one of the browser's trusted certificate authority.

   ANSWER:
   CORRECT, unsigned certificate makes it difficult to enforce the authenticity of the server, not the confidentiality of the communication.

9. For any XML document, the XPath expression "`//drink`" always returns the same result as the expression "`/*//drink`".

   ANSWER:
   INCORRECT, the ‘‘/*//drink’’ would *NOT* return the ‘‘drink’’ element if it is the root element, but ‘‘//drink’’ would.

10. Among the four software layers of a typical Web site, the transport encryption layer is easy to scale through the scaling-out approach.

    ANSWER:
    CORRECT, SSL layer is relatively easy to scale out among the four software layers of a typical web site.

11. If the load on a Web site incurs significantly more write requests than read requests on its database, replicating the database to multiple machines is a good choice for dealing with the growing load.

    ANSWER:
    INCORRECT, the replication is only good for handling READ requests, while the partitioning is good for handling WRITE requests.

12. A Web site that removes all HTML tags from the user input is unlikely to suffer from the cross-site request forgery.

    ANSWER:
    INCORRECT, cross-site request forgery is not related to the user input validation.

13. In a public-key infrastructure, a user can sign an important message with his/her public key in order to guarantee the authenticity and the integrity of the message.

ANSWER:

`INCORRECT, the user should sign the message with his/her *PRIVATE* key to guarantee the authenticity and the integrity of the message.`

14. We use the KMP string-matching algorithm to match a string to a document. In this case, once the ith character in the document starts being compared against a character in the string, no earlier characters in the document (i.e., the characters at the location 1 through $i-1$ in the document) would be compared again.

ANSWER:

`CORRECT, the value of the variable` m `in the KMP algorithm in our note never decreases.`

15. Because of the same-domain policy for Cookies, it is not possible for a malicious hacker who controls the domain A to obtain the content of a cookie from a different domain B through a cross-site request-forgery attack.

ANSWER:

`CORRECT, the same-domain policy disallows a site from reading a cookie from another domain.`

Problem 2: 20 points

Each correct answer to this problem will give you 5 points. There is no penalty for any incorrect answer.

1. In the space provided below, briefly explain how StackGuard detects and prevents a buffer-overflow attack in no more than 100 words.

   ANSWER:

   The StackGuard would add a random number at the stack boundary (right next to the return address after the function call), so that if a buffer-overflow can be easily detected by checking whether the random number stays the same at the end of the function call.

2. Given the following XML document, how many DOM nodes will be produced according to the W3C Document Object Model? (Here, ■ indicates the end of the document.)

   ```
   <class><cs144>You are <em>almost</em> done!</cs144></class>■
   ```

   ANSWER:

   6.

3. You operate a Web server that returns book information given a client request. Now, assume that your server just received a request for the following two books:

   ```
   ISBN: 1234, Title: First book, Author: Josh
   ISBN: 2346, Title: Second book, Edition: 1
   ```

   Your server uses the JSON format to transmit data. Write the JSON-formatted reply from your server that contains the information of the two books. Your answer should be succinct yet self explanatory.

   ANSWER:

   ```
   [{ISBN: 1234, Title:  "First book", Author:  "Josh"},
   {ISBN: 2346, Title:  "Second book", Edition:  1}]
   ```

4. You need to write a Javascript code that gradually increases the font of the text “It is over!!!” from the initial size of 12 pixel to 1000 pixel. The increase should happen at the rate of 1 pixel per 100 ms and should stop once the font size reaches 1000 pixel.

   Complete the Javascript below by adding your code to the empty boxes. Make your code as succinct as possible. It is OK to leave some of the boxes empty.

```
<html>
<head>
  <script type="text/javascript">
    function f(x) {
      var t = document.getElementById("banner");
      var s = parseInt(t.style.fontSize) + parseInt(x);
      t.style.fontSize = String(s) + "px";









    }
  </script>
</head>

<body                                                    >

  <div id="banner" style="font-size: 12px;">It is over!!!</div>
</body>
</html>
```

ANSWER:

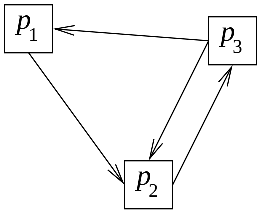
```
In the first box:
if (s < 1000) setTimeout("f("+x+")", 100);
or
if (s < 1000) setTimeout("f(1)", 100);

In the second box:
onLoad="f(1);" (or onLoad="setTimeout(’f(1)’, 100);")
```

Problem 3: 10 points

Again, each correct answer to this problem will give you 5 points without any penalty for an incorrect answer.

Consider the Web of the following three pages:



1. Assuming the damping factor (or random jump probability) is $\frac{1}{4}$, write down the PageRank equation of the three pages, p_1, p_2, and p_3 in the *matrix form.*
   ANSWER:
   $$\begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \frac{3}{4} \begin{bmatrix} 0 & 0 & \frac{1}{2} \\ 1 & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix} * \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} + \frac{1}{4} \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$
   `Note: it's okay for students to use either` p_1 `or` $PR(p_1)$`.`

2. We decide to compute the PageRank values iteratively through the PageRank equation, starting from the initial PageRank values of $(p_1, p_2, p_3) = (1/3, 1/3, 1/3)$. Write the PageRank values of three pages after *one* iteration.

   ANSWER:
   $$\begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} 5/24 \\ 11/24 \\ 8/24 \end{bmatrix}$$

Problem 4: 10 points

You are the administrator of a Web site that hosts a Web discussion forum. You need to design a plan for accommodating the peak load on your site. Your site uses the scaling-out approach to handle the user traffic.

It is known that one DBMS server on a single machine can handle the maximum of 100 READ or WRITE requests per second. For example, one DBMS machine can be configured to handle 90 READ requests/sec and 10 WRITE requests/sec or to handle 20 READ requests/sec and 80 WRITE requests/sec. You can assume that you can store practically infinite data on a single machine.

Your site will host a *single* discussion forum, where all postings to the forum will be stored as tuples in a single table. It is expected that the user activities on the forum will generate 200 READ requests/sec and 50 WRITE requests/sec during the peak. Each READ or WRITE request accesses multiple tuples in the table (often more than 20 tuples). The tuples accessed in each READ/WRITE request is random. That is, it is difficult to predict what tuples will be accessed together in a single request.

Given this setting, answer the following questions. Each correct answer to this problem will give you 5 points without any penalty for an incorrect answer.

1. Briefly explain how you would configure DBMS on multiple servers to deal with the expected peak load on the site. Note: You must justify why you made such a choice.
   Will you replicate your database? YES/NO (please circle)
   Why? ____
   ____

   Will you partition your database? YES/NO (please circle)
   Why? ____
   ____

   ANSWER:
   `YES (Replication). Replication is necessary because one machine cannot deal with 200 READ requests/sec. When we replicate the table to multiple machines, READ requests can be distributed to those machines.`

   `No (Partition). Because WRITE requests access unpredictable tuples, WRITE requests cannot be easily distributed through partitioning.`

2. What is the minimum number of DBMS machines that are needed to handle the expected peak load? ______

Briefly explain the expected number of READ/WRITE requests/sec on each machine for this configuration.

ANSWER:

```
The minimum number of machines is 4.  Each server handles 50 READ and 50
WRITE requests.  That is:  S1(R50,W50), S2(R50,W50), S3(R50,W50), S4(R50,W50).
```