MIEIC   DISTRIBUTED SYSTEMS – 10/11   FEUP

Exam type: cheat sheet   2nd **Semester Exam**
Duration: 2 hours   **21 of June of 2011**
Total points: 20 points
Exam structure: Part I (multiple choice, 50%); Part II (theoretical, 50%).
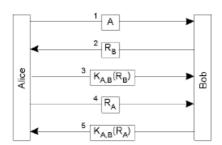
Name:_____
Student ID:_____

Part I: Multiple choice questions [10 points]

Note: For each question there is **only one** correct answer; circle the one in this sheet.

Grading: each correct answer is awarded 1 point; each wrong answer is awarded –0.3 points; non-answered questions, ambiguous, or unintelligible are awarded 0 points. Total of 10 points.

1. A client executes a *remote procedure call* (RPC). Even though the server executes the request, it fails to send a reply. What should the client do if the system uses the semantic "at most once"?
   a. Continues the execution after sending the request
   b. Sends a new request if a time-out occurs
   c. Sends a new request to confirm that the previous one was executed
   d. Remains blocked, and eventually resumes if a time-out occurs

2. A distributed system uses the Berkeley algorithm to sync the local clocks of 3 machines in the system. The time-daemon starts a sync process at 2:05:00. If the clocks of the other two machines are 2:04:50 e 2:05:22, what will be the time in each machine after the synchronization?
   a. 2:05:00; 2:05:00; 2:05:00
   b. 2:05:04; 2:05:04; 2:05:04
   c. 2:05:06; 2:05:06; 2:05:06
   d. 2:05:22; 2:05:22; 2:00:22

3. In a system that uses a *Lamport Logical Clock*, which of the following is correct
   a. The processes use the Lamport algorithm to update the local clocks
   b. The processes increment the clock before sending an event, and append the value of the clock to the message to be sent
   c. The processes increment the clock in a specific interval and update the value of the clock when they receive a message with higher time
   d. A process only updates its clock when it receives a message with time higher than the local one

4. In a distributed file system, it is preferable to provide a remote access model because
   a. The server is not a potential point of congestion of the system
   b. Facilitates consistency maintenance
   c. It is possible to use an automatic mounter to maintain consistency
   d. It is possible to use state-aware servers

5. *Network file system* (NFS) handles RPC failures using
   a. *Dynamic Object Invocation*
   b. Caching in the server and IDs in the RPC transactions
   c. Events and notifications services
   d. *Asynchronous Method Invocation*

6. What makes a failure model be byzantine
   a. Intermittent failures
   b. Arbitrary behavior after a failure occurs
   c. The failure results in a crash of the system
   d. The servers stop answering requests after the failure, although the clients work normally

7. With respect to the Google File System (GFS) discussed in the lectures: Which is the correct answer
   a. The files are split in *chunks* of 64MBytes
   a. *Chunks* are stored in *chunk servers*
   b. Clients interact with the *GFS master* to obtain the *chunk server* that shares the *chunk* the client wants to access to, and with the *chunk server* to read/write in a chunk
   c. All answers are correct

8. When a name is resolved in a given namespace, iterative is worse than recursive resolution because
   a. Communication between machines that are far apart are not costly
   b. There is no way to use the caching system of the server in an efficient way
   c. Requires more work from the name servers
   d. Name resolution of the client is simpler

9. Consider the following authentication protocol, where R_A and R_B are "*once-in-a-lifetime*" *(nonce)* numbers. Which of the following sentences is true?
   a. The protocol is vulnerable to reflection attacks
   b. The protocol is vulnerable to attacks by repeating old messages
   c. The protocol is based on symmetric keys
   d. The protocol is vulnerable because the first message is not encrypted

10. In the properties ACID, which of the following sentences if not valid
   a. *Atomic*: if a transaction cannot execute all operations, some of the operations will still be visible if executed in the correct order
   b. *Consistent*: transactions do not change the integrity of the data structures
   c. *Isolated*: if two or more transactions are executing at the same time, the final result is the same as executing the transactions sequentially
   d. *Durable*: In case of commit, the effects of a transaction are permanent

Part II: Predominantly theoretical [10 valores]

Note: Briefly justify your answers ; **Grading**: Shown in the questions.

1. **[1 pt]** Give a brief definition for *distributed system*, and provide 4 key features to its proper functioning.

2. **[5 pts]** Suppose a file transfer application that allows a client to transfer files to and from the server
   a. Enumerate and briefly describe **three** basic operations that should be supported by this service. For each operation specify its arguments and return values.
   b. Assume you would have to implement such application using the internet protocol stack. In your opinion, which transport protocol is appropriate? Why?
   c. Yet another option to use directly Internet's transport protocols would be to use asynchronous communication. What are the main advantages (at least 2) and disadvantages (at least 2) of such an implementation?
   d. Suppose you would like to provide efficient and reliable access to data. What kind of distributed file system would you recommend for the server? Why (justify your answer discussing the advantages over other distributed file systems)?
   e. Consider concurrency in the server side. Explain its main advantage and how it could be implemented to explore such advantage. Are there any race conditions risks associated with this implementation?
   f. Suppose you would like to implement access control. One possibility is to authenticate users using the (*username*, *password*) pair. Describe **one** potential security problem, and how you could solve it.
   g. Identify one potential failure in the server (other than concurrency). Explain its consequences and how could you minimize them (**Hint**: discuss replication and consistency).

3. **[2 pts]** About names in distributed systems
   a. Explain the appropriateness of naming services in distributed systems.
   b. Considering its structure, names can be classified in two types. Describe them.
   c. Suppose the domain *fa.up.pt* is not a zone, but instead it is integrated in the *up.pt* zone. Describe the resolution process of the name *www.fa.up.pt* from a computer in the subdomain *eslab.upc.edu* (assuming it is a zone). Assume that the A entry associated with *www.fa.up.pt* is not cached. (**Hint**: draw a figure)

4. **[2 pts]** About *Distributed Hash Tables*
   a. Using the Chord Algorithm, discussed in the lectures, calculate the *finger tables* of the following nodes 0, 1, 2, e 6 (**Hint**: draw a figure). Consider an address space between 0 and 8 and that only these nodes are connected.
   b. Knowing that node 0 has item 7 and node 1 item 1, which are the visited nodes to answer *query(7)* (i.e., where is item 7) in node 1?
   c. Enumerate **two** advantages and **one** disadvantages of the Chord algorithm.