NTNU
Norges teknisk-naturvitenskapelige universitet

Fakultet for informasjonsteknologi, matematikk og elektroteknikk

Institutt for datateknikk og informasjonsvitenskap

TDT4190 Distributed Systems
5. june, 2010
0900 – 1300

- **Contact during exam:**
  Svein-Olaf Hvasshovd, tlf. 90049802.

Support:
D - No printed or handwritten supports allowed. Dedicated, simple calculator allowed

Question 1, Time and Global State (15%)

a. Should physical of logical clocks be used if you are interested in ordering the sequence of events in a distributed system? Argue the choice.
b. Explain how the Berkley-algorithm for internal synchronization of physical clocks works.
c. Assume two events e1 and e2 with respectively logical times (Lamport timestamp) k1 and k2. Operator "=>" indicates the relation "happened before". Which of the following assumptions are correct?
   i) e1 => e2 implies k1 < k2.
   ii) k1 < k2 implies e1 => e2.
   iii) k1 < k2 implies e2 NOT=> e1.

Question 2, Coordination and Agreement (15%)

a. Two alternative algorithms for distributed mutual exclusion are "central server" and "ring based". Which one of these is the most fault tolerant? Explain and argue the choice.
b. Assume reliable multicast based on IP multicast. How to make sure the receiver gets all messages?
c. Assume four byzantine generals. One general is a traitor. Can they obtain agreement? Argue your choice.

Question 3, Fault Tolerance (40%)

a. How can it be detected that a server has failed in a distributed multi-node system? Explain and argue your choice.
b. Explain and describe the VP-protocol ("virtual partition protocol"). Emphasize which phases of the protocol.

c. How can the VP-protocol be used to avoid that two parts of a multi-node system continues to operate independently of each other after a network partition?
d. How can a system be prepared for multi instance faults of a node within a short time span?
e. Explain the method used in ClustRa for transfer of responsibility and tasks ("take-over") from a failed node to an operative stand-by node ("hot stand-by node").
f. Explain and describe the "gossip architecture".
g. Assume n servers each with an independent probability p of failure. How can the probability of the availability of a stored object be calculated?
h. What is required of an availability class 5 system?

Question 4, Data Base System (30%)

Your task is to design a data base system that meets the following requirements:

i. Not unavailable more than 5 minutes per year.
ii. Responds within 4 milliseconds for at least 95% when performing single tuple update transactions.
iii. Can scale from 10000 transactions per second to at least 1000000 transactions per second.

Explain in particular the following aspects in your reply:

a. Overall node-architecture.
b. Data distribution architecture.
c. Handling of node faults.