

Sirindhorn International Institute of Technology Thammasat University

Final Examination: Semester 2/2007

Course Title : ITS 413 Internet Technologies and Applications

Instructor : Dr Steven Gordon

Date/Time : Thursday 6 March 2008, 13:30 – 16:30

Instructions:

- ③ This examination paper has __ pages (including this page).
- ③ Condition of Examination
Closed book (No dictionary, No calculator allowed)
- ③ Students are not allowed to be out of the exam room during examination. Going to the restroom may result in score deduction.
- ③ Turn off all communication devices (mobile phone etc.) and leave them under your seat.
- ③ Write your name, student ID, section, and seat number clearly on the answer sheet.
- ③ The space on the back of each page can be used if necessary.

Part A – Multiple Choice Questions [15 marks]

Select the most accurate answer (only select one answer). Each correct answer is worth 1.5 marks. You receive 0 marks for an incorrect answer or no answer.

1. The advantages of a peer-to-peer system (versus a client/server architecture) include:
 - a. Load-sharing, controlling access to resources, and simple algorithms
 - b. Load-sharing, fault-tolerance, and scalability**
 - c. Load-sharing, scalability, and controlling access to resources
 - d. Controlling access to resources, scalability, and fault-tolerance
 - e. Controlling access to resources, simple algorithms, and fault-tolerance
 - f. Scalability, simple algorithms and fault-tolerance

2. The following protocol/system uses a client/server architecture for storing the entire resource index, but a peer-to-peer architecture when peers access resources:
 - a. Gnutella
 - b. Chord
 - c. Jabber/XMPP
 - d. AOL Instant Messenger
 - e. Fasttrack
 - f. Yahoo Messenger
 - g. Napster**

3. What is the approximate per-user throughput for an IEEE 802.11g wireless LAN which contains 9 clients communicating via a single access point (select the closest answer):
 - a. 540Mb/s
 - b. 54Mb/s
 - c. 5.4Mb/s
 - d. 250Mb/s
 - e. 25Mb/s
 - f. 2.5Mb/s**
 - g. 250kb/s

4. The following system is designed to support the network layer handover of a group of nodes at the same time.
 - a. NEMO**
 - b. Mobile IP
 - c. Wireless LAN
 - d. MANET
 - e. None of the above

5. If IPsec in transport mode is used to encrypt a message from a Computer A to a Computer B, then:
 - a. An intermediate router that receives a packet from A to B can view the contents of the message.
 - b. Intermediate routers that receive a packet from A to B cannot view the contents of the message.**

- c. The local router that Computer A is connected to can view the contents of the message.
 - d. The local router that Computer B is connected to can view the contents of the message.
 - e. An intermediate router that receives a packet from A to B cannot determine that A is communicating with B.
6. The main aim of TOR is to:
- a. Hide the contents of the messages between a source and destination user
 - b. Hide the identity of the source and destination from other users**
 - c. Hide the identity of the source from the destination
 - d. Hide the identity of the destination from the source
 - e. Provide a Virtual Private Network between users
 - f. None of the above
7. Which instant messaging protocols use asymmetric servers:
- a. Jabber
 - b. AIM and MSN**
 - c. Yahoo Messenger and Jabber
 - d. AIM
 - e. Yahoo Messenger
 - f. XMPP and Jabber
 - g. XMPP and MSN
8. What delivery method is best suited to Video on Demand delivery in IPTV networks:
- a. Routing
 - b. Broadcast
 - c. Multicast
 - d. Unicast**
 - e. File-based
 - f. Non-real-time
9. What scheme can be used to reduce the impact of hidden terminals in a wireless LAN:
- a. DCF
 - b. PCF
 - c. RTS/CTS**
 - d. Basic Access
 - e. Decreasing the Contention Window
 - f. None of the above
10. A problem with (or disadvantage of) Mobile IP is:
- a. Mobile nodes must change their IP addresses, causing TCP connections to break
 - b. Normal Internet routing will not work when a mobile node changes networks
 - c. A foreign network cannot support visiting mobile nodes that have different Home Agents
 - d. A mobile node must inform the Home Agent of all possible Foreign Agents it will visit before Mobile IP will work
 - e. Some packets must be sent on a sub-optimal route via the Home Agent**
 - f. None of the above

Part B – General Questions [85 marks]

Question 1 [6 marks]

- a) Explain the purpose of a robot exclusion file and how it works (including where the file is and what information the file contains, and how it controls robots – but *you do not have to give the format of the file*). [3 marks]

Answer:

A robot exclusion file is intended to let web robots (or crawlers) know which parts of a website they should access. The file is “robots.txt” and is located inside a directory of the web server (usually the root directory, e.g. www.example.com/robots.txt). The file contains a set of directives as to what files/directories a robot can access (allow) or not (disallow). The directives can be targeted to specific robots (e.g. Googlebot) or all robots.

When a compliant robot accesses the web server, it first checks for the existence of the robots.txt file. If exists, the robot follows the directives, that is, only access the allowed files/directories.

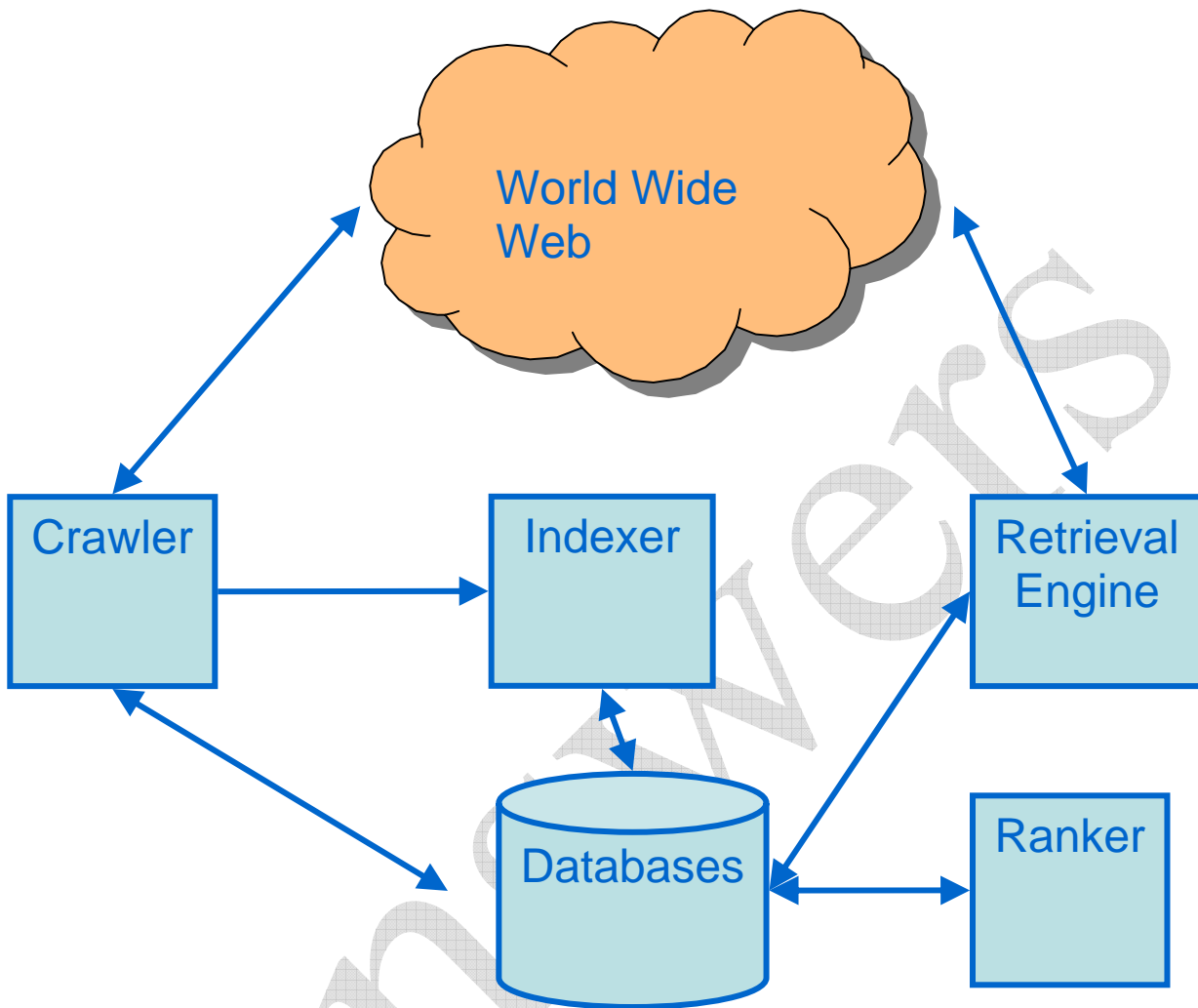
- b) Is a robot exclusion file suitable for protecting (that is, restricting access to) content on a web site? Explain your answer [3 marks]

Answer:

No. A non-compliant robot (that is, a robot that does not adhere to the robots exclusion protocol) will not follow the directives and can then access the files. Also, web browsers usually do not adhere to the robots standard, and hence can still access the files.

Question 2 [21 marks]

- a) Draw the architecture of a typical search engine. Make sure you label each component and show the connectivity between components. [6 marks]



- b) Explain how a search engine crawler works. [4 marks]

Answer:

A crawler is initialized with a set of URLs to visit. The crawler visits a URL, one-at-a-time, and retrieves the page and extracts the links from within the page. The links that have not yet be visited are added to a "To be crawled" list. Those that have been visited are stored in a "Already crawled" list. The crawler then selects a URL from the "to be crawled" list for the next page to visit, and repeats the process.

- c) A search engine crawler picks a URL from a "To be crawled" list. Why is the picking algorithm important? [2 marks]

Answer:

Because a crawler cannot usually traverse all pages on the web (because too many to index within a reasonable amount of time, and the pages change over time), the crawler must select the page to visit that provides must useful content.

- d) Explain the difference between query dependant and query independent ranking algorithms. Give an example set of criteria for each and also give an advantage of each approach (compared to the other approach). [4 marks]

Answer:

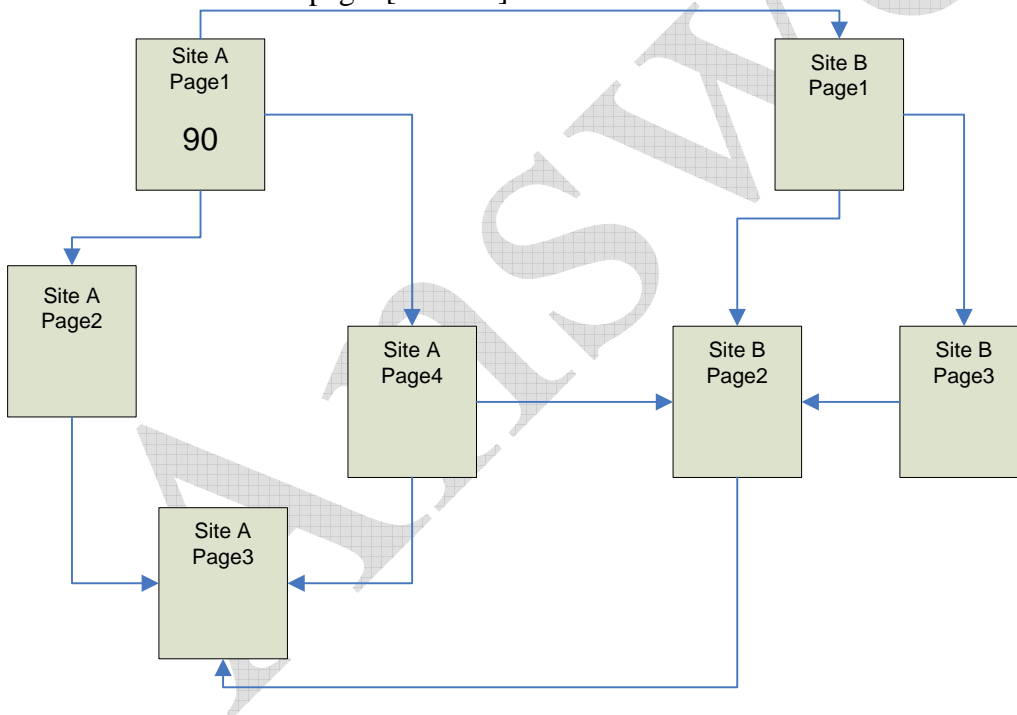
Query dependent ranks pages based in the search criteria/query supplied by the user, whereas query independent ranks pages independent of the user query.

Query dependant example: count of query terms in page; closeness of terms in page

Query independent example: number of links to page; from page; number of accesses

Advantage of independent is that the rank can be pre-calculated (before the user submits a query).
 Advantage of dependant is that is gives results matching the users query.

- e) Calculate the Google PageRank for each page in the diagram below. Write the PageRank value inside the page. [3 marks]



Answer:

- Site A Page 2: 30
- Site A Page 3: 90
- Site A Page 4: 30
- Site B Page 1: 30
- Site B Page 2: 45
- Site B Page 3: 15

- f) Explain a limitation of the Google PageRank algorithm. [2 marks]

Answer:

New sites, even those with very useful information, will initially have a low rank because few (if any) existing sites will link to them. As a result of a low rank they will appear lower in search results, meaning few people will visit them. Hence few people will find out about the new site, meaning very hard to get others to link to it.

Answers

Question 3 [12 marks]

- a) Explain how the Gnutella Peer-to-Peer protocol works, including:
- How do nodes join the network
 - How does a node search the network for a resource

You *do not* have to give details of the protocol, a brief textual description is sufficient. [4 marks]

Answer:

Initially a node must be pre-configured with at least one other peer when it wants to join the network. It then sends a Ping message to this peer who may respond with a Pong, indicating it is willing to act as a permanent peer. The Ping may also be forward to other peers (in a broadcast fashion). When enough Pongs are received, the new node selects C peers as permanent peers.

When a node searches for a resource it sends a query message to its permanent peers. Each peer forwards the query to its permanent peer in a broadcast fashion until the resource is found. When the resource is found, the nodes returns the result to the initiating node, along the same path as the request/query came.

- b) Using the diagram below (which shows a set of nodes and their C=3 permanent peers), answer the following questions (assume the nodes have already joined the network and a node forwarding a message counts as one transmission):
- i. How many times are messages sent in the network using the normal Gnutella protocol if node 1 searches for a resource that is located on nodes 11 and 18 (assume TTL=7)? [2 marks]
 - ii. How long does it take for node 1 to receive a reply from part (i) (assume the hop time is 100 milliseconds)? [2 marks]
 - iii. How many replies are received by node 1? Explain your answer. [2 marks]

In all parts you must explain any additional assumptions you make.

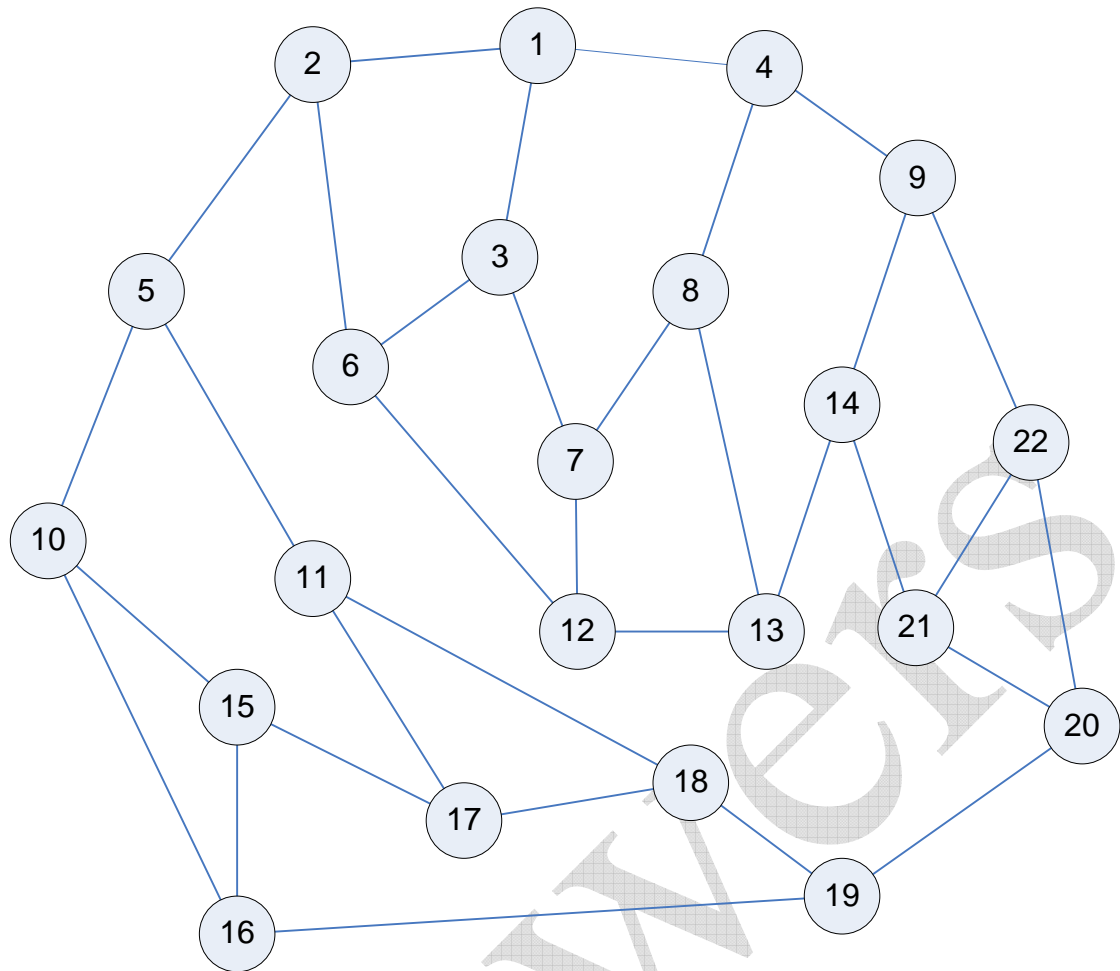
Answers:

Part i. 41 messages. This assumes when two nodes have a message to transmit to each other, that one will transmit first and hence the other will refrain from sending. Also transmissions occur in rounds. There are 32 messages of sending queries and node 11 sends a response (3 hops) and node 18 sends a response (6 hops).

If we assumed that if two nodes were to transmit at the same time, then there would be an extra 5 messages: total 46.

Part ii. $6 \times 100 = 600\text{ms}$

Part iii. 2. A response from node 11 and a response from node 18.



- c) How does the Time To Live (TTL) assist, with respect to broadcast overhead, in Gnutella?
[2 marks]

Answers:

TTL limits the number of times a packet is forwarded so that the packet isn't forward forever (especially if loops exist).

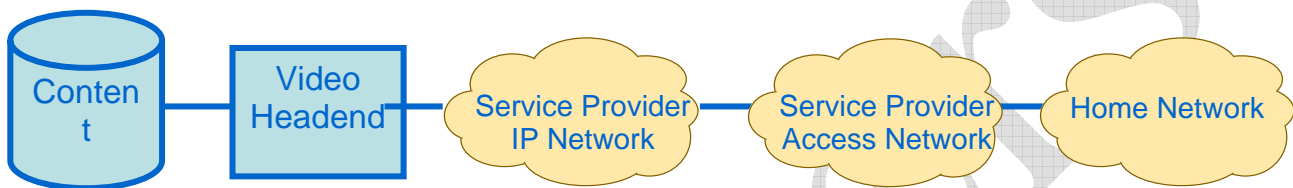
Question 4 [15 marks]

- a) Explain the difference between IPTV, File-based TV distribution and Internet (or Net) TV. You should also mention the network (or type of network) that each are delivered over. [2 marks]

Answer:

Net TV is the real-time delivery of low to medium quality TV over the Internet (usually via web sites). File based TV is non-real-time delivery of entire content over Internet (usually P2P file sharing). IPTV is delivery of high quality TV (equivalent to digital TV) over IP based network (usually over private IP network).

- b) The figure below shows a typical structure of an IPTV network.



- i. Which portion of the network is the performance bottleneck? [1 mark]
- ii. List two network technologies that may be suitable for delivering IPTV into a home (that is, the technology used to connect your home to a larger service provider network). You must give accurate names of standards/protocols, referring to versions if necessary. [2 marks]
- iii. Fibre-to-the-X may be used to refer to a range of different fibre-based technologies, where X can have different names.
 - List two possible Fibre-to-the-X technologies for IPTV [1 mark],
 - Describe briefly the two technologies [2 marks], and
 - Explain the trade-offs (advantages and disadvantages) between the two technologies [3 marks]
- iv. For the delivery of HDTV channels to users over the Service Provider IP Network, the bandwidth requirement of the network may be measured in terms of number of channels delivered. However, for the delivery of Video on Demand to users, the bandwidth requirement of the Service Provider IP Network may be measured in terms of number of VoD subscribers. Explain why the bandwidth requirement is measured with different metrics for the different services (TV versus VoD). [4 marks]

Answers:

i. Service Provider Access Network

ii. ADSL2, ADSL2+, IEEE 802.11g/a/n, Ethernet, Optical Fibre or FTTH, WiMax

iii. Fibre-to-the-Home: deliver optical fibre directly from the Service Provider IP network to the home, bypassing the need for copper cables. Fibre-to-the-Curb: deliver optical fibre to the street level (covering 10's of homes), and then use copper to deliver from a street-level cabinet to each home; Fibre-to-the-Node: deliver optical fibre to a neighborhood level exchange (covering 100's of

homes), and then use copper to deliver to each home. The tradeoffs are cost of deployment (the more use of optical fibre, the high the cost, mainly for installation – hence FTTH is most expensive, FTTN is least expensive) versus performance (copper wires only allows relatively slow speeds compared to optical fibre – hence FTTH performs the best, FTTN is the slowest).

iv. For normal TV delivery, content is sent at scheduled times to many users, and hence multicast can be used. Therefore, approximately, each channel will require a constant bandwidth requirement across the IP network, independent of the number of users. That is, one copy of the channel will be sent across the core of the network, while multiple copies sent at the edge of the network (when delivering to individual users). For VoD, each individual user effectively has their own tailored content (own movie and their own time), and hence multicast cannot be used. A copy of each each video must be sent for each user, hence the bandwidth requirement depends on the number of users.

ANSWERS

Question 5 [10 marks]

An average performance comparison of the Napster, Fasttrack and Chord P2P systems is given below:

	Latency (delay) in search	Messages to be sent for search	Update Cost	Storage Required
Napster	1	1	1	1
Gnutella	$\log(n)$	n	1	1
Fasttrack	$\log(C)$	C	1	1
Chord	$\log(n)$	$\log(n)$	$\log(n)$	$\log(n)$

where C is the number of super-peers and n is the number of nodes in the network. Note that the values represent an “order of magnitude” – they don’t represent absolute values.

Assume a P2P network last year had 10,000 nodes. However, rapid growth in the popularity of the network saw the size grow to 1,000,000 nodes. (In the cases where super-peers are used, assume C is 1% of the total number of nodes). Answer the following questions. You may use this example network to help explain your answers.

- Explain why the “Messages to be sent” for Gnutella is in the order of n . [2.5 marks]
- Explain why the “Update cost” for Napster is in the order of 1 . [2.5 marks]
- With respect to only the performance criteria shown in the table, which P2P system is the best for the example network. Explain your answer. [2.5 marks].
- Considering the performance criteria in the table AND the advantage of fault-tolerance, which P2P system is the best for the example network. Explain your answer. [2.5 marks]

Answers:

a. For Gnutella, messages are broadcast to nodes when searching. Therefore, with more nodes, more messages need to be broadcast in the search. The number of messages is proportional to the number of nodes. Therefore, increasing the number of nodes by 100 (from 10,000 to 1,000,000), will see the number of messages increased by 100.

b. An update involves a node adding a resource. For Napster, a node adds a resource by informing the central server. The update is independent of the number of nodes. Hence, the cost is constant.

c. Napster. All performance criteria are constant (1), compared to the Chord where all performance criteria are 6.

d. Chord. All performance criteria are doubled (from 3 to 6), but still quite small (Napster is 1, whereas Gnutella/Fasttrack are 1,000,000 and 10,000 for messages). But Chord has the advantage over Napster of fault-tolerance – it does not rely on a central server.

Question 6 [21 marks]

Chord is a protocol that uses Distributed Hash Tables for peer-to-peer applications.

- a) Draw a diagram of an example Chord network that has 8 nodes (and can support no more than 8 nodes) [2 marks]

Answer:

The diagram will be a circle with 8 points on it, numbered in order, 0 through to 7.

- b) Using your example network where necessary, explain:
- How are nodes given identifiers that represent their position in the Chord network? [2 marks]
 - What is the relationship between keys and resources in Chord? [1 mark]
 - What is the relationship between keys and nodes? [2 marks]
 - What other nodes does a node maintain routes to? [2 marks]
 - In addition to the addresses of other nodes, the routing information maintained by a node should also contain what? [2 mark]

Answers:

Part i. An address of a node (e.g. IP address) is hashed to determine an ID of a node.

Part ii. A resource identifier is hashed to determine a resource key.

Part iii. A key is stored on the node with the same value of ID. If the node does not exist, then the key is stored on the next node in the ring that does exist.

Part iv. A node maintains routes to nodes that are 2^n positions away, e.g. 1, 2, 4, 8, In the example, node 1 maintains routes to nodes 2, 3, 5. Node 2 maintains routes to nodes 3, 4, 7.

Part v. The keys stored by that node (and as a consequence, the keys that node is responsible to routing to).

Assume nodes 3, 4 and 7 have left your example network (that is, there are only 5 nodes remaining).

- c) For each node, list the keys the node stores. [2.5 marks]

Answer:

Node 1: 1
Node 2: 2
Node 5: 3, 4, 5
Node 6: 6
Node 0: 0, 7

- d) For each node, list the other nodes the node maintains routes to. [2.5 marks]

Answer:

Node 1: 2, 5
Node 2: 5, 6
Node 5: 6, 0, 1
Node 6: 0, 2
Node 0: 1, 2, 5

e) Describe the path taken if node 2 sends a query for resource with key 1. [2 marks]

Answer:

Node 2 sends the query to node 6. Node 6 sends the query to node 0. Node 0 sends the query to node 1.

f) If Chord was modified such that each node maintained routes to every second node, then explain an advantage and disadvantage compared to the actual Chord protocol. [3 marks]

Answer:

Advantage: Search would be faster, because a node knows about more nodes (and hence more keys) in the network

Disadvantage: Route maintenance would be much higher, because a node must maintain routes with $n/2$ nodes