Problem Set #4

Problem 1 Hashing (20pts)

The hash table has 13 slots, and integer keys are hashed into the table with the following hash function H

```
int H (int key)
{
        x = ( key + 5 ) * ( key - 3 );
        x = int( x / 7 ) + key;
        x = x % 13;
        return x;
}
```

(a) Fill in the final hash table with the following keys: 17, 22, 73, 56, 310, 100, 230, 12, 42, 18, 19, 24, 49
10 points (reduce points based on how many values are wrong.)

Slot	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Contents	**49**		**18**		**22,42**			**310,12,24**	**73,100,19**	**17**	**56**	**230**	

(b) List 2 methods that can handle collision in hashing. 10 points – 5 each method
If the key and values are stored together in the table. When lookup returns multiple results, an additional lookup of the key value can always identify the correct value.

Separate Chaining and Open Chaining are 2 methods using the above mentioned techniques.
Other methods demonstrating reasonable collision management ability also counts. Since Hashing is trying to use extreme small space to store large amount of data and collision is inevitable down the road.

Problem 2 Distributed Hash Tables (20pts) 5 for each table

There is a Chord DHT in Figure 1. with 5 nodes. The finger tables are listed beside the nodes. Each node may be storing some items according to the Chord rules (Chord assigns keys to nodes in the same way as consistent hashing)

(a) Fill in the table for node id=1 and 7

	ID+2^i	successor
0	2	3
1	3	3
2	5	6

Table for ID=1

	ID+2^i	successor
0	0	1
1	1	1
2	3	3

Table for ID=7

(b) List the node(s) that will receive a query from node 1 for item 5 (item named by key 5) **10 points if all correct**

It will be Node 1 -> Node 6.

Node 1 will find the largest node id which does not exceed the item you are looking for(which is 5). Node 1 will find the last row in its finger table(Node 5) and goes to the successor(Node 6)

And Node6 indeed have the item, the query is concluded.

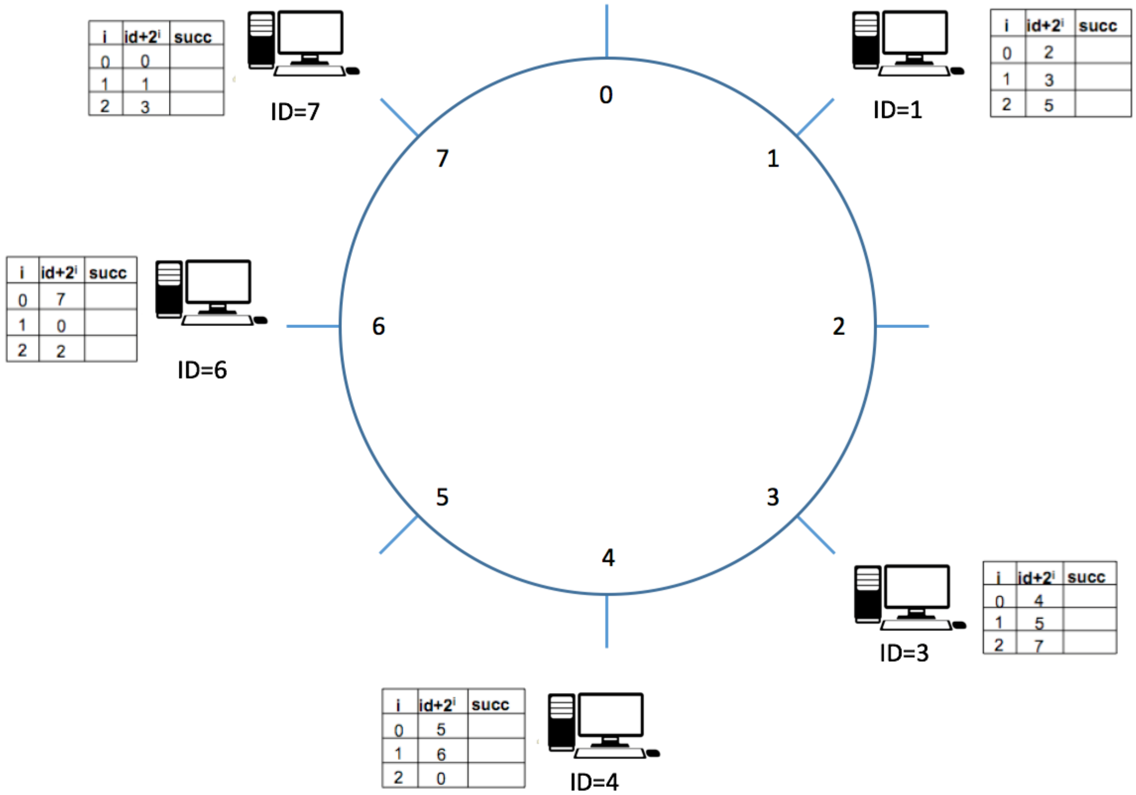

Figure 1. Chord DHT for Problem 2

Problem 3 Bloom Filters (10pt)

Derive the probability of false positive rate after 20 keys (or elements) are inserted into a table of size 100. Assume that 4 hash functions are used to setup bit positions in the table for the keys (elements).

4 Hashing functions means: For each key (element) there will be 4 bits to be set to 1 in the table (it can also be less than 4 if the hash functions generate same outputs).

Hint:

Assume **m** is the number of bits in the filter, **n** is the number of elements and k is the number of hash functions used.

After inserting one key, the probability of a particular bit being 0 is $(1-\frac{1}{m})^k$. This is because k hash functions are independent, and each hash function will have $(1-\frac{1}{m})$ probability for a particular bit remain 0. Then after inserting n keys, the probability for a particular bit remain 0 is $(1-\frac{1}{m})^{kn}$

Now you have to apply this to the case of false positive.

Sol:

Assume **m** is the number of bits in the filter, **n** is the number of elements and k is the number of hash functions used.

After inserting one key, the probability of a particular bit being 0 is (1-1/m)^k. This is because k hash functions are independent, and each hash function will have probability for a particular bit remain 0. Then after inserting n keys, the probability for a particular bit remain 0 is (1-1/m)^(kn)

False position means the bloom filter believe an element is inserted but in reality it is not.

It means for an element that haven't been inserted, the probability of k hashing functions all point to bits that are already set to 1 is the probability of false positive.

P = (1-(1-1/m)^(kn))^k

If we let m=100, n=10 and k=5 we get P=0.0096

The answer is 0.0096

Problem 4 P2P system (10pt)5 points for each subquestion

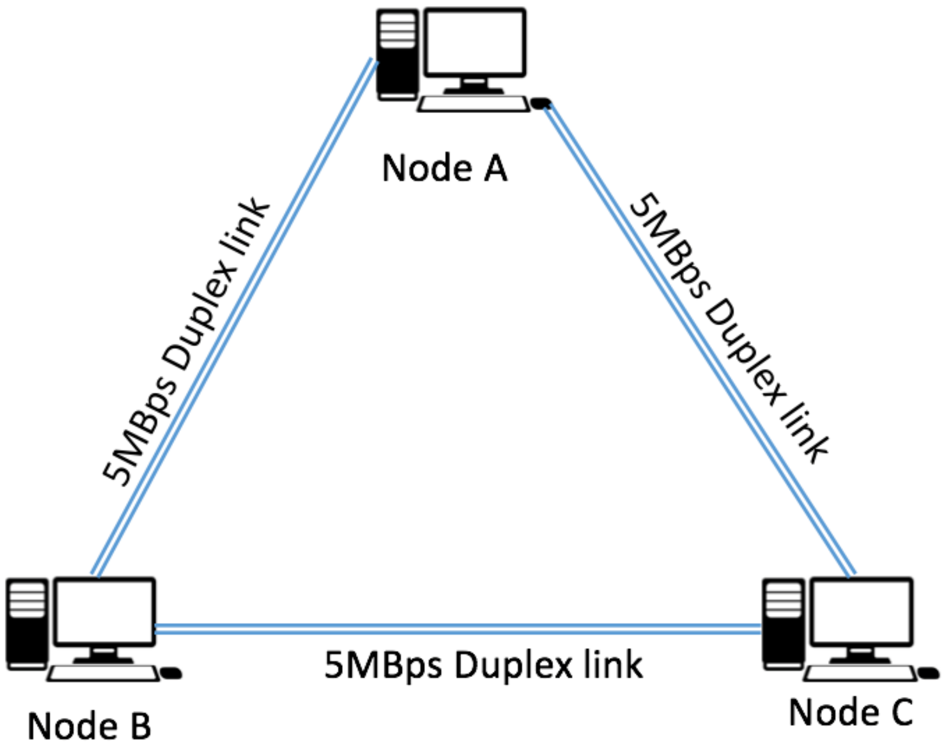

Figure 2. Network topology for Problem 4

3 nodes A, B and C are connected with each other via 5MBps duplex link as is now in Figure 2. Node A want to share a 2500MB file to Node B and C. During the actual transmission 2.5MB piece of the file can be send on the links each time. In the problem we ignore the RTT delay.

(a) What is the time of the sharing process using centralized approach? (The process ends when B and C all received the file, and the centralized approach means B and C are communicating with A independently and there is no communication between B and C)
The finish time will be max(2500MB/5 MBps, 2500MB/5MBps)
= 500s

(b) What is the ideal minimum time of the sharing process using P2P approach? (P2P approach means after B and C received a piece from A, they immediately share the their piece to other party)
The ideal scenario is that B and C received different half of the whole file, and while receving from A, they constantly sharing what they received with each other.

At time 250s, B has received 500pieces from A and 499piceces from C, this is because C just received the 500th piece itself and haven't have time to send it to B yet. And same situation is for C.
B and C need to exchange the last piece and it takes 2.5MB/5MBps = 0.5s
The total time is 250.5s.
Answer of 250 sec is also considered correct.

Problem 5 File Distribution (20pt) 10 for each table or 2.5 each entry

Consider distributing a file of F = 20 Gbits to N peers. The server has an upload rate of us = 30 Mbps, and each peer has a download rate of di = 2 Mbps and an upload rate of u. For N = 10 and 100 and u = 300 Kbps and 2 Mbps, prepare a chart giving the minimum distribution time for each of the combination of N and u for both client-server distribution and P2P distribution.

Solution

For calculating the minimum distribution time for client-server distribution, we use the following formula:

$$D_{cs} = max\ \{NF/u_s,\ F/d_{min}\}$$

Similarly, for calculating the minimum distribution time for P2P distribution, we use the following formula:

$$D_{P2P} = max\{F/u_s,\ F/d_{min},\ NF/(u_s + \sum_{i=1}^{N} u_i)\}$$

Where, F = 20 Gbits = 20 * 1024 Mbits u_s = 30 Mbps d_{min} = d_i = 2 Mbps

Client Server

u	N= 10	N= 100
300 Kbps	10,240	68,267
2 Mbps	10,240	68,267

Peer to Peer

u	N= 10	N= 100
300 Kbps	10,240	34,536
2 Mbps	10,240	10,240

Problem 5 BGP (20pt)

1. Give the types of business relationships in BGP peering and mention who pays whom. What conditions make the Internet stable? Explain each condition in a line or two.

Customer-Provider (Customers pay provider for access to the Internet)
Peer-Peer (Peers exchange traffic free of charge) 4 points(2 each)

1.Route export 6 points (2 each)
– Don't export routes learned from a peer or provider to another peer or provider
2. Global topology
– Provider-customer relationship graph is acyclic E.g., my customer's customer is not my provider
3. Route selection
– Prefer routes through customers over routes through peers and providers

2.

Please identify which of the following paths are valid, which of them are invalid based on the network topology of Figure 3. 10 points (roughly 1 each)

Path 1 3 d invalid
Path 1 4 d valid
Path 8 d valid
Path 6 d valid
Path 4 d valid
Path 7 5 d valid
Path 7 5 3 d valid
Path 2 1 3 d invalid
Path 1 4 6 d valid

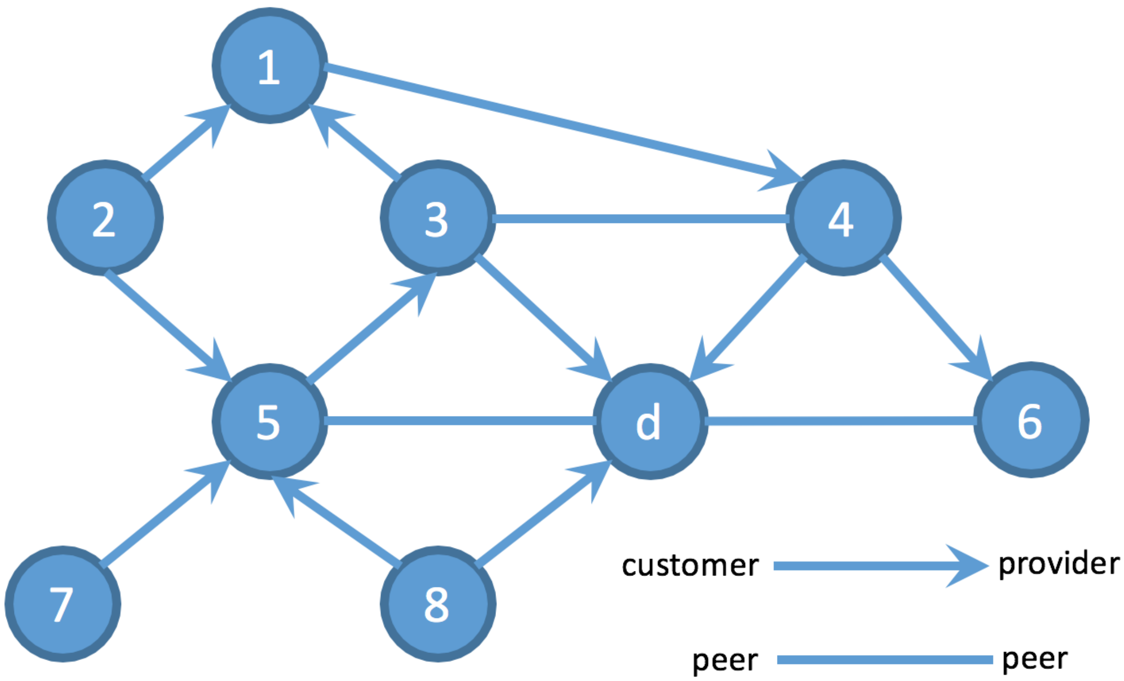

Figure 3. Network topology for Problem 5