Gnutella

(Group no. 5)

A. Gnutella is an open, decentralized, P2P search protocol that is mainly used to find and share files

B. Peer-to-Peer networks have become a major research topic over the last few years. Object location is a major part in the operation of these distributed systems.

C. The original Gnutella algorithm uses flooding (BFS traversal of the underlying graph) for object discovery and contacts all accessible nodes within the TTL value. Although it is simple and manages to discover the maximum number of objects in that region, the approach does not scale, producing huge overhead to large numbers of peers

D. Due to these disadvantages , some other search techniques have been proposed

E. In Random Walks, the requesting node sends out k query messages to an equal number of randomly chosen neighbors. Each of these messages follows its own path, having intermediate nodes forward it to a randomly chosen neighbor at each step. These queries are also known as walkers.

F. Types of Techniques :

1. Flooding
   (a) TTL(d) where d is depth of search.(TTL is time to live)
   (b) if r is the average degree of the graph then the no. of terms traversed is $r + r^2 + r^3 + ... + r^d$

2. Random Walk
   (a) probability of finding an object ->p. The object is residing in p fraction of the node.
      i. probability that random number will succeed in first hop is p
      ii. probability that it will succeed in second hop is $(1-p)p$
      iii. probability that it will succeed in second hop is $(1-p)^2p$
      iv. probability that it will succeed in Tth hop is $(1-p)^{T-1}p$

(b) So, Expected number of hops $X = 1.p + 2(1-p)p.... + T(1-p)^{T-1}p +$ ......Solving the equation is (1/p).

(c) K random walk

   i. 1st hop - kp
   ii. 2nd hop - $(1-p)^{k}kp$
   iii. 3rd hop - $(1-p)^{2k}kp$
   iv. Tth hop - $(1-p)^{(T-1)}kp$

(d) So, Expected number of hops $X = 1.kp + 2(1-p)^{k}kp.... + T(1-p)^{(T-1)k}kp +$ ......Solving the equation is (1/kp).

(e) Biased Random Walk →

   Each node records the degree of the neighboring nodes. Search easily gravitates towards high degree nodes that hold more clues.

3. Dyanmic Querying

- Expanding Ring
   i. Start wih TTL(1)
   ii. Wait
   iii. See how many results obtained
   iv. If not desired no. of results obtained go for TTL = 2

   D = Total no. of $\vec{d}$esired results
- r = no. of results received
   R = no. of results needed or $(D - r)$
   H = no. of peers already queried
   RH = r/H = result density (Note : RH is not equal to R * H)
- R/RH = HQ = Estimate of the number and peers required
- d = to run for TTL(d)
- $X^{d} = HQ$
- Average degree of the network = Y
   therefore, $d^{th} hop \approx Y^{d}$
   $Y > HQ$
   where $X^{d} \approx HQ$

4 . One hop replication

- Each node keeps track of the indices of the files belonging to its immediate neighbors.
- As a result, high capacity / high degree nodes can provide useful clues to a large number of search queries

5. Two layered network architecture :

- A structure similar to Kazaa can also be used.
   and store indices of the nodes which are under them .
- In this method some nodes act like "super" peers