The cross-ISP traffic and performance tradeoff in VoD system

May 10, 2011

VoD System Model

- There are M ISPs: ISP 1, ISP 2, ISP 3,..., ISP M.
- The peer number and average peer upload bandwidth in different ISPs:
  - the peer number: We use the ON-OFF model. There are totally N_i peers (who have installed the software for VoD streaming) in ISP i. A part of N_i peers stays offline. The probability that a peer stays offline is π_0. The probability that there are $N_i^{off} = m_i^0$ peers offline is $P(N_i^{off} = m_i^0) = C_{N_i}^{m_i^0} \pi_0^{m_i^0} (1 - \pi_0)^{N_i - m_i^0}$.
  - The online peers are downloading chunks and uploading chunks in the system.
  - the average peer upload bandwidth in ISP i is U_i.
  - A total of J constant-length chunks to be shared in the VoD system: $C_1, C_2, ..., C_J$.
  - The cache of a peer can store B chunks, $B << J$.
- The time is slotted. In a time slot T, peers play one chunk. The probability that a peer has cached the chunk it is playing is small. We assume peers need to download at least one chunk in a time slot.

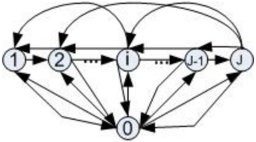
Chunk Demand in ISP i

- A total of J constant-length chunks to be shared in the VoD system: $C_1, C_2, ..., C_J$.
- At time slot T, peers download the chunks that is playing. So, the chunks should be downloaded in less than the time slot interval δT.
- In ISP i, at time slot T, there are m_i^j peers downloading chunk j, and there are m_i^0 offline peers. $m_i^0 + m_i^1 + ... + m_i^J = N_i$. Say peers are in state $0, 1, 2, ..., J$ as peers are offline, downloading chunk 1, downloading chunk 2,...,downloading chunk J. The probability that peers are in some state is π_j for state j. $\sum_{j=0}^{J} \pi_j = 1$.
- The probability that there are $m_i^j (0 \leq j \leq J)$ peers in state j is $P(m_i^0, m_i^1, ..., m_i^J) = N_i! \frac{\pi_0^{m_i^0}}{m_i^0!} ... \frac{\pi_J^{m_i^J}}{m_i^J!}$.
- The probability that there are m_i^j peers downloading chunk $j (1 \leq j \leq J)$ is $P(m_i^j) = C_{N_i}^{m_i^j} \pi_j^{m_i^j} (1 - \pi_j)^{N_i - m_i^j}$.

Chunk Demand in ISP i

- At a time slot, every peer in state j send requests for chunk j. So, the number of requests, k is the same as the number of peers in state j, m_i^j. It is a random variable of binomial distribution, $P(Req = k) = C_{N_i}^k \pi_j^k (1 - \pi_j)^{N_i - k}$. For large N_i, the binomial distribution can be approximated by the Poisson distribution, $P(Req = k) = \frac{\lambda_{i,j}^k}{k!} e^{-\lambda_{i,j}}$, $\lambda_{i,j} = N_i \times \pi_j$. So, the requests for chunk j in a time slot is a random variable of Poisson distribution, $P(Req = k) = \frac{\lambda_{i,j}^k}{k!} e^{-\lambda_{i,j}}$, $\lambda_{i,j} = N_i \times \pi_j$.
- Request rate for chunk 1,...J are $\lambda_{i,1}$,...,$\lambda_{i,J}$ respectively.

The chunk popularity in VoD system

transition for user behavior.jpg 

- User behaviors can be modeled by the state transition of peers. Based on the state transition model for user behavior, we can get the stationary state distribution for peer state, $(\pi_0, \pi_1, ..., \pi_{J-1}, \pi_J)$. We can get the chunk popularity from this.(This is the same user behavior model proposed by Yipeng Zhou in an infocom2011 paper)
- User behaviors: Joining, Departures, Random seek.

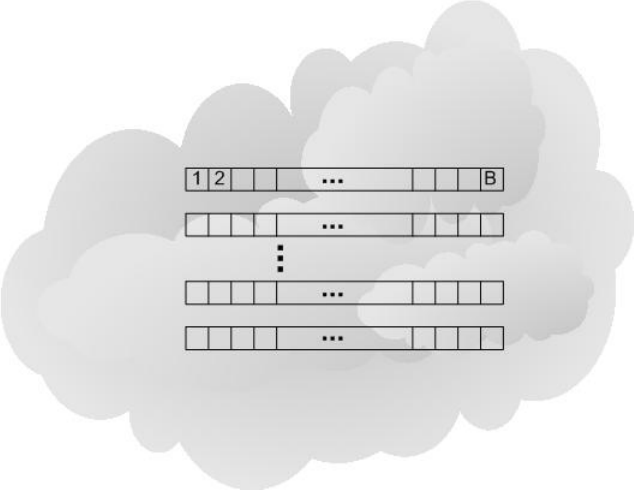
Define performance metrics

- The resource used to serve chunk j is from peers or from servers.
- At a time slot, m_i^j peers demand chunk j, we assume no peers playing the chunk in the cache(the probability is small). let w_j denote the copies of chunk j that online peers can upload.
- The needed server capacity to satisfy the demand for chunk j is $U_{sj} = max\{m_i^j - w_j, 0\}$. The total needed server capacity is $U_s = \sum_{j=1}^{J} U_{sj}$.
- If the server capacity is given as S, the probability of chunk missing for streaming is $P = \frac{\sum_{j=1}^{J} U_{sj} - S}{N_i - m_i^0}$.

chunk distribution in peers' cache

- The cache of a peer can store B chunks, $B < J$.
- We assume the cache of all peers are filled up at beginning according to replication strategy in P2P VoD.
- We use the proportional replication strategy as the replication strategy. The chunk popularity for chunk j is $p_j = \frac{\pi_j}{1-\pi_0}$. The number of replicas of chunk j in ISP i is $n_i^j = p_j N_i B$.

chunk distribution in peers' cache



Chunk requests and service

- Considering ISP i with no inter-ISP links, requests for chunk j in ISP i are uniformly directed to peers who has cached chunk j in ISP i. The average request rate for chunk j received by a peer caching chunk j is $\frac{\lambda_{i,j}}{n_i^j} = \frac{1-\pi_0}{B}$. As there are B chunks stored in a peer's cache, the total request rate received by a peer is $1 - \pi_0$.
- The peer upload bandwidth is equally divided into B parts to serve the requests for B different chunks stored in the peer respectively.

Locality

- We assume every peer keeps x inter-ISP neighbors. A peer sends the request for chunk j to the peers having cached chunk j in other $N_i - 1$ peers in the same ISP and x inter-ISP neighbors.
- In the $N_i - 1$ peers in the same ISP and x inter-ISP neighbors, there are $p_j \cdot (N_i + x)B$ peers having cached chunk j in average(the peer which sends requests for chunk j doesn't cache chunk j).
- Requests for chunk j are uniformly directed to peers who has cached chunk j. With the x inter-ISP neighbors, the request rate for chunk j from peers in ISP i transmitting through inter-ISP links to other ISPs is $\frac{\lambda_{i,j}}{p_j \cdot (N_i + x)B} \cdot (p_j \cdot x \cdot B) = \frac{x}{N_i + x} \lambda_{i,j}$.
- Requests for chunk j to ISP i from other ISPs: the request rate for chunk j from ISP k to ISP i is $\frac{N_k}{N - N_i} \frac{x}{N_i + x} \lambda_{k,j}$. So, the requests rate for chunk j from other ISPs to ISP i is $\sum_{k=1,k!=i}^{k=M} \frac{N_k}{N - N_i} \frac{x}{N_i + x} \lambda_{k,j}$.

Next step work

- Calculate the system performance based on the above model.
- Derive the relationship between system performance and inter-ISP neighbor size x.