# Draft of VoD modeling

Jian Zhao

## Modeling of Traffic Locality in VoD Streaming Systems

The primary objective is to determine the performance of VoD streaming systems when locality-aware peer selection strategy is applied in the VoD streaming systems. The locality-aware peer selection strategy is that peers have the limited number of neighbors who are in other ISPs. In this paper, we intend to use differential equations to derive the needed downloading time for video chunks demanded by peers.

### *System Model*

There are $K$ ISPs. The number of peers in ISP $k$ is $n_k$. The peer arrival rate in ISP $k$ is $\lambda_k$.

A movie is composed of $M$ chunks. Chunk is used to advertising to neighbors what parts of a movie a peer holds. The size of a chunk is $L$. The streaming rate is $R$. So the playback time for a chunk is $\frac{L}{R}$.

At one moment, every peer is downloading a specific chunk. We denote $n_{0,m}$ as the number of peers who is downloading chunk $m$, $1 \leq m \leq M$. When there are no user interactions, the peer after finishing downloading chunk $m$ will download chunk $m+1$.

### *Performance*

Calculate the downloading time for different video chunks demanded by peers under two situations: locality-unaware peer selection and locality-aware peer selection.

*locality-unaware peer selection strategy:* When locality-unaware peer selection strategy is taken in the VoD system, the number of peers downloading chunk $m$ in all ISPs is $n^m$. We don't consider the user interactions first. We can use the following differential equations to describe the change of number of peers downloading different chunks:

$$\frac{dn_{0,1}}{dt} = \lambda - \gamma_{0,1} n_{0,1}$$

$$\frac{dn_{0,m}}{dt} = \gamma_{0,m-1} n_{0,m-1} - \gamma_{0,m} n_{0,m}, 2 \leq m \leq M$$

$$\gamma_{0,M} = \frac{R}{L}$$

$$\gamma_{0,M-1} = \frac{U_p \cdot n_{0,M}}{(n_{0,1} + \ldots + n_{0,M-1})L}$$

$$\gamma_{0,m} = \gamma_{0,m+1} + \frac{U_p \cdot n_{0,m+1}}{(n_{0,1} + \ldots + n_{0,m})L}, 1 \leq m \leq M-2$$

In this case, the inter-ISP traffic is:

$$T_1 =$$

*locality-aware peer selection strategy:*

$$\frac{dn_{k,1}}{dt} = \lambda_k - \gamma_{k,1} n_{k,1}$$

$$\frac{dn_{k,m}}{dt} = \gamma_{k,m-1} n_{k,m-1} - \gamma_{k,m} n_{k,m}, 2 \leq m \leq M$$

$$\begin{aligned}
\gamma_{k,M} &= \frac{R}{L} \\
\gamma_{k,M-1} &= \frac{U_p \cdot n_{k,M} \cdot p}{n_{k,1} + \ldots + n_{k,M-1}} \\
&+ \sum_{i=1, i\neq k}^{M} \frac{U_p \cdot n_{i,M} \cdot (1-p)}{(n_1 + \ldots + n_{M-1}) - (n_{i,1} + \ldots + n_{i,M-1})} \\
\gamma_{k,m} &= \gamma_{k,m+1} + \frac{U_p \cdot n_{k,m+1} \cdot p}{n_{k,1} + \ldots + n_{k,m}} \\
&+ \sum_{i=1, i\neq k}^{M} \frac{U_p \cdot n_{i,m+1} \cdot (1-p)}{(n_1 + \ldots + n_m) - (n_{i,1} + \ldots + n_{i,m})}
\end{aligned}$$

In this case, the inter-ISP traffic is:

$$T_2 =$$

The key points are how to calculate the average downloading rate of a specific chunk. One way to calculate the average downloading rate of a specific chunk is based on the statement that the total downloading rate of all downloaders is equal to the total upload rate from uploaders.

The second way to calculate the average downloading rate of a specific chunk is to calculate the probability that a peer connects to neighbors that can upload chunks to it.

The next step is to add the model of user interactions in the differential equations.

As I think over the calculation of average downloading rate for the above differential equations, the asymmetric downloading characteristic of VoD system (younger peers download chunks from older peers) make me consider how this characteristic can influence the total needed server capacity to support peers' downloading rate and the peers' remaining upload bandwidth that can support new arriving peers. I have done the following work about it.

### *the theoretical system capacity*

In this section, I try to model the change of the random variables: the needed server capacity and the peers' remaining upload bandwidth that can support new arriving peers, as the peer number in the system increases. We assume there are two kinds of peers in the system with upload bandwidth $U_1$,$U_2$ respectively. We denote $S_n$ as the needed server capacity as there are $n$ peers in the system, denote $B_n$ as the peers' remaining upload bandwidth as there are $n$ peers in the system.

$$B_n = \begin{cases} B_{n-1} - R + X_n & \text{if} B_{n-1} \geq R \\ X_n & \text{if} B_{n-1} < R \end{cases}$$

$X_n$ is the upload bandwidth of the new arriving peer when there are $n-1$ peers in the system. The above equations can be written as:

$$B_n = X_n + \max(0, B_{n-1} - R); B_1 = X_1$$

So, $B_n$ is only related to the state of one previous step,$B_{n-1}$. The random process $B_n$ is a Markov Chain.

We can analyze the above Markov Chain $B_n$ under three cases: 1)$U_1 + U_2 \geq 2R$; 2)$U_1 + U_2 = 2R$; 3)$U_1 + U_2 \leq 2R$. The analysis of the case $U_1 + U_2 = 2R$ is easier because the states of $B_n$ under this case are not complicated. We can try to get the analytical solutions for $B_n$. With $B_n$, we can calculate the total needed server capacity $S_n$:

$$S_n = S_{n-1} + \max(0, R - B_{n-1})$$

An approximate solution for large $n$ may be obtained.

$$\begin{aligned} B_n = X_n + \max(0, B_{n-1} - R) \geq X_n + B_{n-1} - R \\ B_n - B_{n-1} \geq X_n - R \\ B_n - B_1 = X_n + X_{n-1} + ... + X_2 - (n-1)R \\ B_n = X_n + X_{n-1} + ... + X_2 + X_1 - (n-1)R \end{aligned}$$

As $n$ is large, we can use the central limit theorem. $\sum_{i=1}^{n} X_i$ can be approximated by a normal distribution. So, when $n$ is large, we can determine when a new peer enters the VoD system, the probability that the new peer needs to contact servers for downloading can be calculated. And the needed server upload bandwidth is:

$$Y_{n+1} = max(0, R - B_n)$$

$B_n$ can be estimated by a normal distribution.

I am thinking over how to analyze the traffic locality among multiple ISPs through this model. There exists one problem in analyzing traffic locality, we need to consider multiple ISPs, so there are different random processes $B_n$, $S_n$ for different ISPs. The difficulty is how to use one random process, which is a Markov process, to model the multiple ISPs with traffic locality.