1 Modeling of the P2P service migration problem

We suppose there are M videos, and N ISPs. There are one on-premise server and one cloud node in each ISP.

1.1 Optimization of the problem without Lyapunov optimization

Notation definition:

C_s^j: storage capacity of the on-premise server at the $j-th$ ISP
C_u^j: upload bandwidth capacity of the on-premise server at the $j-th$ ISP
h_j: charging rate for storage on the cloud at the $j-th$ ISP
k_j: charging rate for upload bandwidth on the cloud at the $j-th$ ISP
s_m: storage of $m-th$ video

$x_m^j = \{0,1\}, m = 1, ..., M$: $x_m^j = 1$ if the placement of the $m-th$ video is on the on-premise server at the $j-th$ ISP; $x_m^j = 0$ otherwise;

$y_m^j = \{0,1\}, m = 1, ..., M$: $y_m^j = 1$ if the placement of the $m-th$ video is on the cloud at the $j-th$ ISP; $y_m^j = 0$ otherwise;

r_m^j: request rate of the $m-th$ video from the $j-th$ ISP, i.e., the bandwidth demand is $s_m r_m^j$.

R_{ji}^m: percentage of requests from j for video m is routed to on-premise server i
T_{ji}^m: percentage of requests from j for video m is routed to cloud i

$min \sum_{m=1}^M \sum_{j=1}^N \sum_{i=1}^N (s_m r_m^j T_{ji} k + s_m h) y_m^j - \alpha \sum_{m=1}^M \sum_{j=1}^N s_m r_m^j (T_{jj} + R_{jj})$ (maximize local traffic, i.e., minimize delay)

subject to:
$y_m^j = \{0,1\}, \forall j = 1, ..., N, \forall m = 1, ...M$
$x_m^j = \{0,1\}, \forall j = 1, ..., N, \forall m = 1, ...M$
$\sum_{i=1}^N (R_{ji}^m + T_{ji}^m) = 1, \forall j = 1, ...N, \forall m = 1, ..., M$
$0 \le R_{ji}^m \le x_m^i, \forall j = 1, .., N, \forall i = 1, ..., N, \forall m = 1, ..., N$
$0 \le T_{ji}^m \le y_m^i, \forall j = 1, .., N, \forall i = 1, ..., N, \forall m = 1, ..., N$
$\sum_{m=1}^M s_m x_m^j \le C_s^j, \forall j$ (on-premise server's storage constraint)
$\sum_{m=1}^M \sum_{j=1}^N s_m r_m^j R_{ji}^m \le C_u^i, \forall i = 1, ..., N$ (on-premise server's upload bandwidth constraint)

Note:
known values: C_s^j, C_u^j, h_j, k_j, s_m, r_m^j
optimization variables: x_m^j, y_m^j, R_{ji}^m, T_{ji}^m

1.2 Optimization of the problem with Lyapunov optimization

This is a combination of optimization for one time deployment and time-average variables. The placement of content is one time deployment while the schedule is for time-average.

Notation definition:

C_s^j: storage capacity of the on-premise server at the $j-th$ ISP

C_u^j: upload bandwidth capacity of the on-premise server at the $j-th$ ISP

h_j: charging rate for storage on the cloud at the $j-th$ ISP

k_j: charging rate for upload bandwidth on the cloud at the $j-th$ ISP

s_m: storage of $m-th$ video

$x_m^j = \{0,1\}, m = 1, ..., M$: $x_m^j = 1$ if the placement of the $m-th$ video is on the on-premise server at the $j-th$ ISP; $x_m^j = 0$ otherwise;

$y_m^j = \{0,1\}, m = 1, ..., M$: $y_m^j = 1$ if the placement of the $m-th$ video is on the cloud at the $j-th$ ISP; $y_m^j = 0$ otherwise;

D_s^{ji} is the delay from source j to on premise server i, and D_c^{ji} is the delay from source j to on cloud node i.

$r_m^j(t)$: at time slot t, number of requests of the $m-th$ video generated from the $j-th$ ISP.

$R_m^{ji}(t)$: at time slot t, number of requests for video m that are routed from region j to on-premise server i

$T_m^{ji}(t)$: at time slot t, number of requests for video m that are routed from region j to cloud node i

$Q_m^j(t)$: at time slot t, queues of requests from video m from ISP j.

Note: The queue update is: $Q_m^j(t+1) = \max[Q_m^j(t) + r_m^j(t) - \sum_{i=1}^N R_m^{ji} - \sum_{i=1}^N T_m^{ji}, 0]$

Different from the previous sub section, $R_m^{ji}(t)$ and $T_m^{ji}(t)$ is not a schedule of fraction of arrival rates for all time slots. Now they are schedule of number of requests (integers) for each time slot.

minimize $k\overline{\sum_{m=1}^M \sum_{j=1}^N \sum_{i=1}^N (s_m T_m^{ji}(t))} + \alpha \sum_{m=1}^M \sum_{j=1}^N \sum_{i=1}^N \overline{s_m R_m^{ji}(t)} + \beta h \sum_{m=1}^M \sum_{j=1}^N (s_m y_m^j) + \gamma \sum_{m=1}^M \sum_{j=1}^N (s_m x_m^j) - \rho \sum_{j=1}^N \sum_{i=1}^N \overline{\sum_{m=1}^M s_m (T_m^{ji}(t) D_c^{ji} + R_m^{ji}(t) D_s^{ji})}$

subject to:

$y_m^j = \{0,1\}, \forall j = 1, ..., N, \forall m = 1, ...M$

$x_m^j = \{0,1\}, \forall j = 1, ..., N, \forall m = 1, ...M$

$0 \le R_m^{ji}(t) \le R_m^{ji}(t) x_m^j, \forall j = 1, .., N, \forall i = 1, ..., N, \forall m = 1, ..., N, \forall t$

$0 \le T_m^{ji}(t) \le T_m^{ji}(t) y_m^t, \forall j = 1, .., N, \forall i = 1, ..., N, \forall m = 1, ..., N, \forall t$

$\sum_{m=1}^M s_m x_m^j \le C_s^j, \forall j$ (on-premise server's storage constraint)

$\sum_{m=1}^M \sum_{j=1}^N s_m R_m^{ji}(t) \le C_u^i, \forall i = 1, ..., N, \forall t$ (on-premise server's upload bandwidth constraint)

Queues $Q_m^j(t)$ is stable, $\forall m, j$, i.e., $\overline{r_m^j(t)} \le \overline{\sum_{i=1}^N R_m^{ji} + \sum_{i=1}^N T_m^{ji}}$

$Q_m^j(0) = 0, \forall m, j$

Note:

known values: C_s^j, C_u^j, h_j, k_j, s_m, $r_m^j(t)$, D_c^{ji}, D_s^{ji}

optimization variables: x_m^j, y_m^j, $R_m^{ji}(t)$, $T_m^{ji}(t)$