

# Appunti utility-based caching

---

## Abstract

Appunti per Utility-based caching. Ho riscritto il modello in maniera comprensibile e togliendo le assunzioni su singola qualità degli oggetti.

*Keywords:* Optimization, Content Distribution.

---

## 1. Modello

Di seguito propongo un modello per utility-based caching. Comincio a descrivere il caso in cui vi sono solamente oggetti di tipo “video”; in seguito aggiungo gli altri come estensione.

Table 1: Summary of the notation used in this paper.

Parameters of the Models	
$A$	Set of arcs
$V$	Set of ASes
$O$	Set of objects
$Q$	Set of qualities
$FS(i)$	Set of forward arcs $(i, j) \in A$ for node $i \in V$
$BS(i)$	Set of backward arcs $(j, i) \in A$ for node $i \in V$
$b_{i,j}$	Capacity of arc $(i, j) \in A$
$n_v^o$	Number of requests for object $o$ , in AS $v \in V$
$r^q$	Rate required to retrieve an object at quality $q \in Q$
$s^q$	Storage space required to cache an object at quality $q \in Q$
$U^q$	Utility gained to serve one request for an object at quality $q$
$p_v^o$	0-1 Producers reachability matrix $p_v^o = 1$ if AS $v$ has a producer for object $o \in O$ (it can serve whatever quality of object $o$ )
$S_{AS}$	Max caching storage that can be installed at an AS
$S_{TOT}$	Max caching storage that can be installed in the network
$bw_v$	Max egress capacity for AS $v \in V$ , $bw_v = \max \left( \sum_{e \in FS(v)} b_e; \sum_{o \in O} n_v^o \cdot \max_{q \in Q} r^q \right)$

Decision Variables of the Models	
$n_v^{o,q}$	Number of requests for object $o$ at quality $q$ satisfied at AS $v$
$x_{v_s}^{o,q}$	0-1 Caching variable, if the source AS $v_s \in V$ caches $o$ at quality $q$
$y_e^{o,q,v_d}$	Flow on arc $e \in A$ for object $o \in O$ , at quality $q$ sent to the destination AS $v_d \in V$
$d^{o,q,v_d}$	Rate requested at AS $v_d \in V$ , for object $o$ at quality $q$
$z_{v_s}^{o,q,v_d}$	Rate provided by the source AS $v_s \in V$ , for object $o$ , at quality $q$ for the destination $v_d \in V$ , when $v_s$ behaves as a producer ( $p_{v_s}^{o,q} = 1$ )
$w_{v_s}^{o,q,v_d}$	Rate provided by the source AS $v_s \in V$ , for object $o$ , at quality $q$ for the destination $v_d \in V$ , when $v_s$ behaves as a cache ( $x_{v_s}^{o,q} = 1$ )

$$\max \sum_{o \in O} \sum_{q \in Q} \sum_{v \in V} n_v^{o,q} U^q \quad (1)$$

subject to:

$$\sum_{q \in Q} n_v^{o,q} = n_v^o \quad \forall o \in O, v \in V \quad (2)$$

$$d^{o,q,v_d} = n_v^{o,q} \cdot r^q \quad \forall o \in O, q \in Q, v_d \in V \quad (3)$$

$$d^{o,q,v_d} = z_{v_d}^{o,q,v_d} + w_{v_d}^{o,q,v_d} + \sum_{e \in BS(v_d)} y_e^{o,q,v_d} - \sum_{e \in FS(v_d)} y_e^{o,q,v_d} \quad \forall o \in O, q \in Q, v_d \in V \quad (4)$$

$$z_{v_s}^{o,q,v_d} + w_{v_s}^{o,q,v_d} + \sum_{e \in BS(v_s)} y_e^{o,q,v_d} = \sum_{e \in FS(v_s)} y_e^{o,q,v_d} \quad \forall o \in O, q \in Q, v_s \in V, v_d \in V, v_s \neq v_d \quad (5)$$

$$\sum_{o \in O} \sum_{q \in Q} \sum_{v_d \in V} y_e^{o,q,v_d} \leq b_e \quad \forall e \in A \quad (6)$$

$$\sum_{v_d \in V} z_{v_s}^{o,q,v_d} \leq p_{v_s}^{o,q} \cdot bw_{v_s} \quad \forall o \in O, q \in Q, v_s \in V \quad (7)$$

$$\sum_{v_d \in V} w_{v_s}^{o,q,v_d} \leq x_{v_s}^{o,q} \cdot bw_{v_s} \quad \forall o \in O, q \in Q, v_s \in V \quad (8)$$

$$\sum_{o \in O} \sum_{q \in Q} x_{v_s}^{o,q} \cdot s^q \leq S_{AS} \quad \forall v_s \in V \quad (9)$$

$$\sum_{o \in O} \sum_{q \in Q} \sum_{v_s \in V} x_{v_s}^{o,q} \cdot s^q \leq S_{TOT} \quad (10)$$

$$x_v^{o,q} \in \{0, 1\} \quad \forall o \in O, q \in Q, v \in V \quad (11)$$

$$n_v^{o,q} \in \mathbb{Z}^+ \quad \forall o \in O, q \in Q, v \in V \quad (12)$$

$$y_e^{o,q,v_d} \in \mathbb{R}^+ \quad \forall o \in O, q \in Q, v_d \in V, e \in A \quad (13)$$

$$d^{o,q,v_d} \in \mathbb{R}^+ \quad \forall o \in O, q \in Q, v_d \in V \quad (14)$$

$$z_{v_s}^{o,q,v_d}, w_{v_s}^{o,q,v_d} \in \mathbb{R}^+ \quad \forall o \in O, q \in Q, v_d \in V, v_s \in V. \quad (15)$$

5 The set of constraints (2) makes sure that all the requests are served at one quality level. In the problem instances we add a “special” quality level  $q = 0$ , which represents unserved traffic demands. When serving quality  $q = 0$

no bandwidth is required:  $r^q = 0$ , moreover, it does not generate any utility  $\forall o \in O, U_r^o = 0$ . The set of constraints (3) set the value of the bandwidth demand at AS  $a$ , for object  $o$ , at quality  $q$ . Such demand is satisfied in (4). In particular, it can be satisfied because: (1) the AS is a producer for that object (i.e.:  $z_{v_d}^{o,q,v_d} = d^{o,q,v_d}$ ), (2) the AS caches the object (i.e.:  $w_{v_d}^{o,q,v_d} = d^{o,q,v_d}$ ), or (3) the AS retrieves the object (i.e.: la somma dei flussi sugli archi entranti).

Flow balance constraints are imposed in (5). We bound the arc capacity in (6); similarly, in (7) and (8), we limit the maximum emitted flows the AS sends when it behaves as a producer and a cache, respectively.

In (9) we restrict the overall caching storage that can be deployed in an AS, and in (10) we set the same limit w.r.t. the entire topology.

E' molto meglio presentare il modello come descritto sopra (con pochi tagli, in modo che il reviewer lo capisca senza troppo sforzo). Si aggiunge poi una section sulle estensioni.

## 2. Estensioni

### 2.1. Non-cooperazione tra gli AS

E' facile aggiungere dei vincoli per prevenire la cooperazione tra gli AS. Per farlo basta cambiare il bilanciamento dei flussi, rimuovere (5), ed aggiungere (16).

$$z_{v_s}^{o,q,v_d} + \sum_{e \in BS(v_s)} y_e^{o,q,v_d} = \sum_{e \in FS(v_s)} y_e^{o,q,v_d} \quad \forall o \in O, q \in Q, v_s \in V, v_d \in V, v_s \neq v_d \quad (16)$$

### 2.2. Low-quality videos

Esistono dei video che possono essere serviti solo a bassa qualità. Sia  $O_{LQ} \subseteq O$  il set di questi oggetti video, low quality. Sia  $q = 1 \in Q$  l'unica qualità a cui possono essere serviti. Questi possono essere facilmente integrati al modello come estensione, aggiungendo i seguenti vincoli:

$$n_v^{o,q} = 0 \quad \forall o \in O_{LQ}, v \in V, q \in Q, q \neq 1 \quad (17)$$

### 2.3. *Max utility, min traffic*

C'è il problema di trovare la soluzione che minimizzi il traffico, fissata l'utilità. Per risolvere questo problema, usare l'approccio multiobiettivo è fattibile, ma è preferibile (principalmente per evitare problemi di instabilità numerica) risolvere il modello in due passi. Nella prima iterazione si calcola il valore della  
35 massima utilità. Nella seconda iterazione si determina la soluzione dell'istanza della rete, minimizzando il traffico, e fissando come utilità il valore massimo trovato al passo precedente. Per ora non lo implementerei e basta.

### 3. Esempi numerici

40      Ho qualche dubbio sul risultato di Fig. 3b nel caso non-cooperative (non avevo il sorgente e magari c'è un refuso nella figura/descrizione). Si dice che i clients di AS 2 scaricano l'oggetto 1 e l'oggetto 2 a qualità 2, ma questo violerebbe i vincoli di banda, dato che servono  $33 \cdot 1500$  Kbps ma ne sono disponibili solamente 20000 Kbps.

45      I risultati sono pressoché equivalenti, anche se la formulazione riportata in questi appunti riesce a calcolare bound di performance migliori evitando di fissare a priori le qualità degli oggetti richiesti, e potendo soddisfare le richieste con qualità differenziate.