Probabilistic Caching Policy for Categorized Contents and Consecutive User Demands

Presenter Minseok Choi

Co-authors: Dongjae Kim, Dong-Jun Han, Joongheon Kim and Jaekyun Moon

Emails: choimins@usc.edu, minseok.choi307@gmail.com

May 22, 2019

Outline

- Basics and motivation
- Content request model
- Network model
- Successful delivery rate for consecutive user demands
- Optimal probabilistic caching policy
- Numerical results

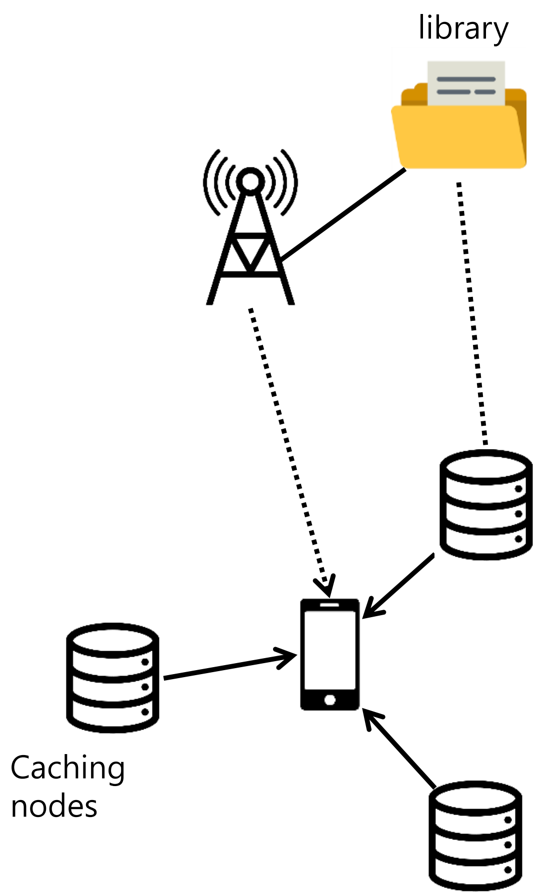
Basics of Wireless Caching Network

- On-demand streaming service
  - A relatively **small number of popular contents** are requested at ultra high rates.
  - **Playback delay** is often the more important measure of goodness to the user than other usual performance metrics like **video quality**.

  → **Wireless caching network** has considered as a promising technique.

- Wireless caching network
  - Bringing popular contents closer to users
  - Independent storages in caching nodes

  → How to cache the content?



Motivation

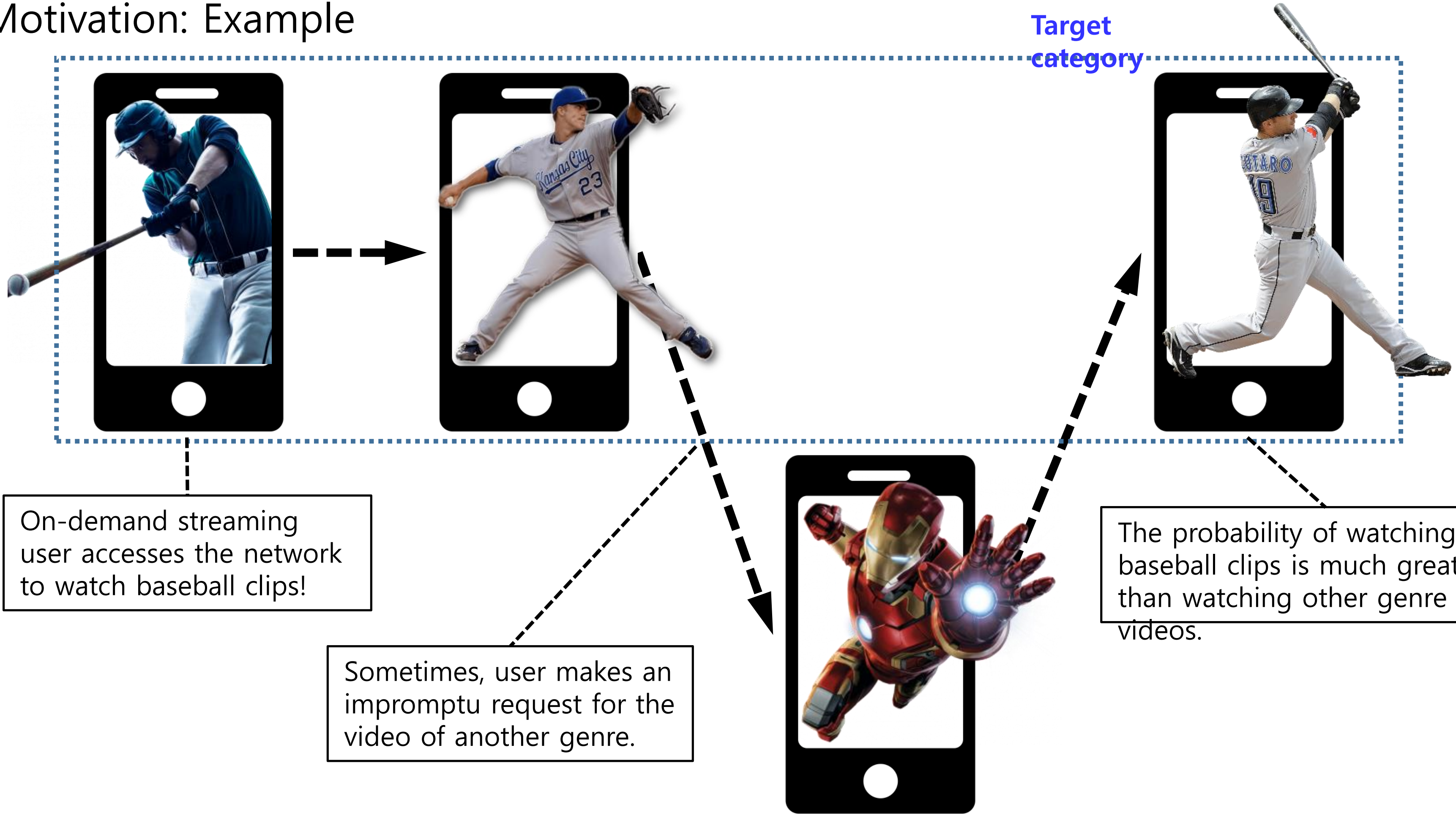
Previous:

- Independent content requests
- Independent content popularity
- Caching policy optimized for one-shot request

Our paper:

- High correlation among multiple consecutive content requests
- Categorized contents
- Caching policy optimized for consecutive requests of categorized contents

Motivation: Example

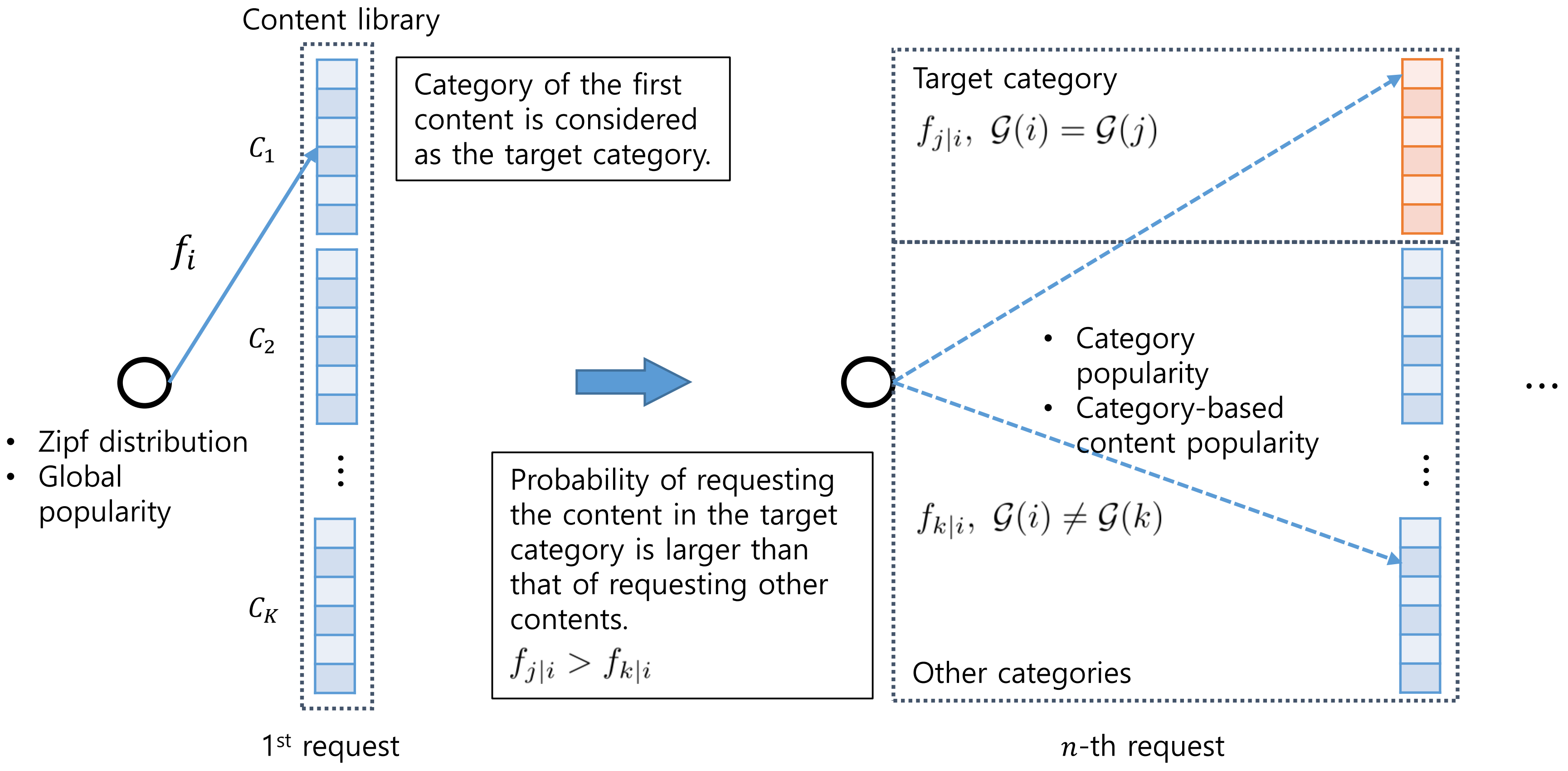


On-demand streaming user accesses the network to watch baseball clips!

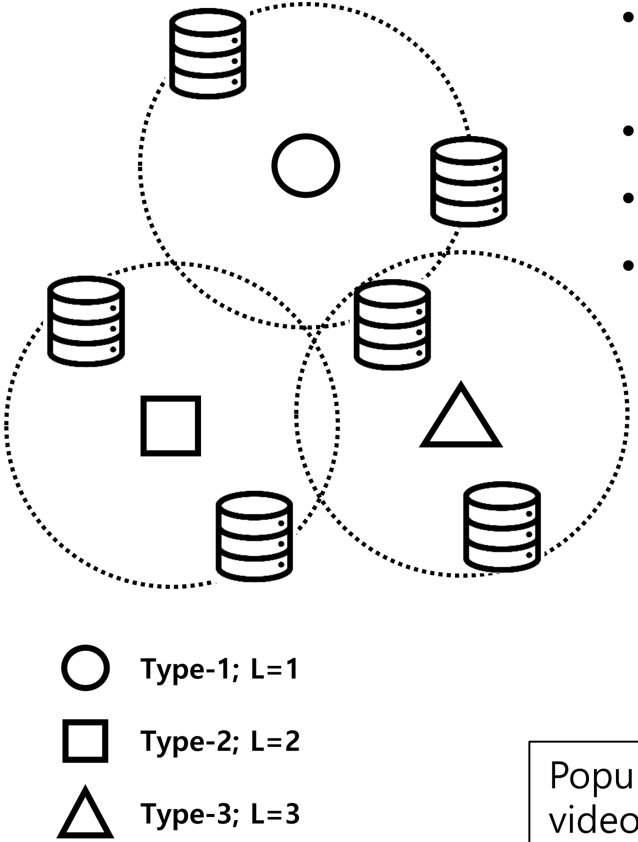
Sometimes, user makes an impromptu request for the video of another genre.

The probability of watching other baseball clips is much greater than watching other genre videos.

Content Request Model



Network Model and Successful Delivery Rate



- Caching nodes are distributed by the independent Poisson point process with intensity of λ.
- The Rayleigh fading channel: $h = \sqrt{D}g$ where $D = 1/d^{\alpha}$
- A type-l user watches consecutively l videos.
- The average successful delivery rate of type-L user:

Data rate for delivering file i from the nearest helper which stores file i.

$$P_L^s = \sum_{i_1} \cdots \sum_{i_L} f_{i_1} f_{i_2|i_1} \cdots f_{i_L|i_1} \cdot \prod_{i=i_1}^{i_L} \Pr\{\tilde{R}_i \geq \rho_L\}$$

$$= \sum_{i_1} \cdots \sum_{i_l} f_{i_1} f_{i_2|i_1} \cdots f_{i_L|i_1} \cdot \boxed{\prod_{i=i_1}^{i_L}} (1 - e^{-Cp_i})$$

Popularity of consecutively consumed videos given file i_1 requested at first.

No repeated request is assumed, i.e., $i_L \neq i_l, \forall l \in \{1, \cdots, L-1\}$.

Optimization Problem Formulation

- Optimization problem to find probabilistic caching policy

$$\mathbf{p}^* = \underset{p_i, i=1,\cdots,F}{\arg\max} \left[\min\{P_1^s, \cdots, P_L^s\} \right] \quad \text{s.t.} \quad \sum_{i=1}^{N} p_i \leq M$$

$$0 \leq p_i \leq 1$$

Lemma 1.

$P_l^o > P_m^o$ for any $l, m \in \{1, \cdots, L\}$ and $l < m$

Lemma 2.

The optimum vector $\mathbf{p}^* = (p_1^*, \cdots, p_F^*)^T$ satisfies $\sum_{i=1}^{F} p_i^* = M$.

$$\mathbf{p}^* = \underset{p_i, i=1,\cdots,F}{\arg\max} P_L^s \quad \text{s.t.} \quad \sum_{i=1}^{N} p_i = M$$

$$0 \leq p_i \leq 1$$

- How to solve?

Dual-Variable Subproblem

- Optimization problem with auxiliary variable
  - Given $\{p_i\}_{i \neq n,m}$, optimize p_m and p_n.
  - The average successful delivery rate of type-L user:

The case in which the user requests both i_m and i_n.

The case in which the user requests i_n but i_m.

$$P_L^s = \overline{a_{m,n}(1-e^{-Cp_m})(1-e^{-Cp_n})} + \underline{b_{m,n}(1-e^{-Cp_m})} + \overline{d_{m,n}(1-e^{-Cp_n})} + \underline{e_{m,n}}$$

The case in which the user requests i_m but i_n.

The case in which the user does not request both files i_m and i_n.

- Dual-variable subproblem

$$\{p_m^*, p_n^*\} = \arg\min_{p_m, p_n} \mathcal{M}_{(p_m, p_n)} \qquad \text{s.t.} \quad p_m + p_n = q_{m,n} = M - \sum_{i=1, i \neq m,n}^{F} p_i$$

where $\mathcal{M}_{(p_m,p_n)} = x_{m,n} e^{-C \cdot p_m} + y_{m,n} e^{-C \cdot p_n}$ $\qquad 0 \le p_m, p_n \le 1$

Solution of Dual-Variable Subproblem

- According to arithmetic-geometric mean inequality,

$$x_{m,n}e^{-C\cdot p_m} + y_{m,n}e^{-C\cdot p_n} \geq 2\sqrt{x_{m,n}y_{m,n}e^{-C\cdot q_{m,n}}}$$

With equality if and only if $x_{m,n}e^{-c\cdot p_m} = y_{m,n}e^{-c\cdot p_n} \iff$

$$\tilde{p}_m = \frac{1}{2C}\log\frac{x_{m,n}}{y_{m,n}} + \frac{1}{2}q_{m,n}$$

$$\tilde{p}_n = \frac{1}{2C}\log\frac{y_{m,n}}{x_{m,n}} + \frac{1}{2}q_{m,n}$$

- **Proposition 1.**

The optimal solution of the dual-variable subproblem is as follows:

$$\{p_m^*, p_n^*\} = \begin{cases} \{\tilde{p}_m, \tilde{p}_n\} & \text{if } 0 \leq \tilde{p}_m, \tilde{p}_n \leq 1 \\ \underset{\{p_m, p_n\}}{\arg\min}\{\mathcal{M}_{(0,q_{m,n})}, \mathcal{M}_{(q_{m,n},0)}\} & \text{elseif } q_{m,n} < 1 \\ \underset{\{p_m, p_n\}}{\arg\min}\{\mathcal{M}_{(1,q_{m,n}-1)}, \mathcal{M}_{(q_{m,n}-1,1)}\} & \text{elseif } q_{m,n} \geq 1 \end{cases}$$

Iterative Algorithm

- To find the optimum $\mathbf{p}$, iteration should be applied with respect to
  $\forall(m,n) \in \{1,\cdots,F\} \times \{1,\cdots,F\},\ m \neq n$
- Convergence
  - The updated objective sequence $\{\mathcal{M}_{(p_m,p_n)}\}$ is nondecreasing.
  - The original objective function is upper bounded, i.e., $P_L^s \leq 1$.

Numerical Results - Parameters

- Category structure

$\mathcal{G}_1$	$\mathcal{G}_2$	$\mathcal{G}_3$	$\mathcal{G}_4$	$\mathcal{G}_5$
1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25

- Popularity model
  - 1st content request (Zipf)

$$f_i = \frac{i^{-\gamma}}{\sum_{j=1}^{F} j^{-\gamma}}, \quad f_i > f_j \ \forall i < j$$

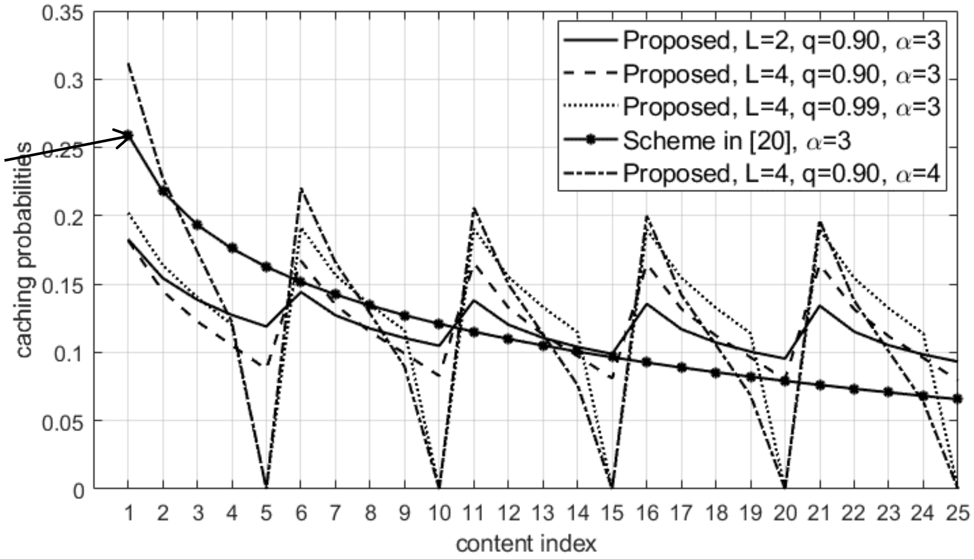
  - Consecutive content requests (Uniform)

$$f_{j|i} = \begin{cases} \frac{q}{(N - \mathcal{N}(\mathcal{G}(i)))} & \text{if } \mathcal{G}(j) = \mathcal{G}(i) \\ \frac{(1-q)}{(F - N - \mathcal{N}(\bigcup_{n \neq i} \mathcal{G}(n)))} & \text{if } \mathcal{G}(j) \neq \mathcal{G}(i) \end{cases}$$

- Maximal number of types: $L = 4$
- Cache size: $M = 3$
- Probability of requesting the preferred category: $q = 0.9$
- The intensity of caching node distribution: $\lambda = 0.2$

Numerical Results – Caching Probabilities

Optimized for one-shot request



L: consecutive content request number
q: probability of requesting contents in the target category
α: pathloss exponent

- Every peak point represents the most popular file in each category

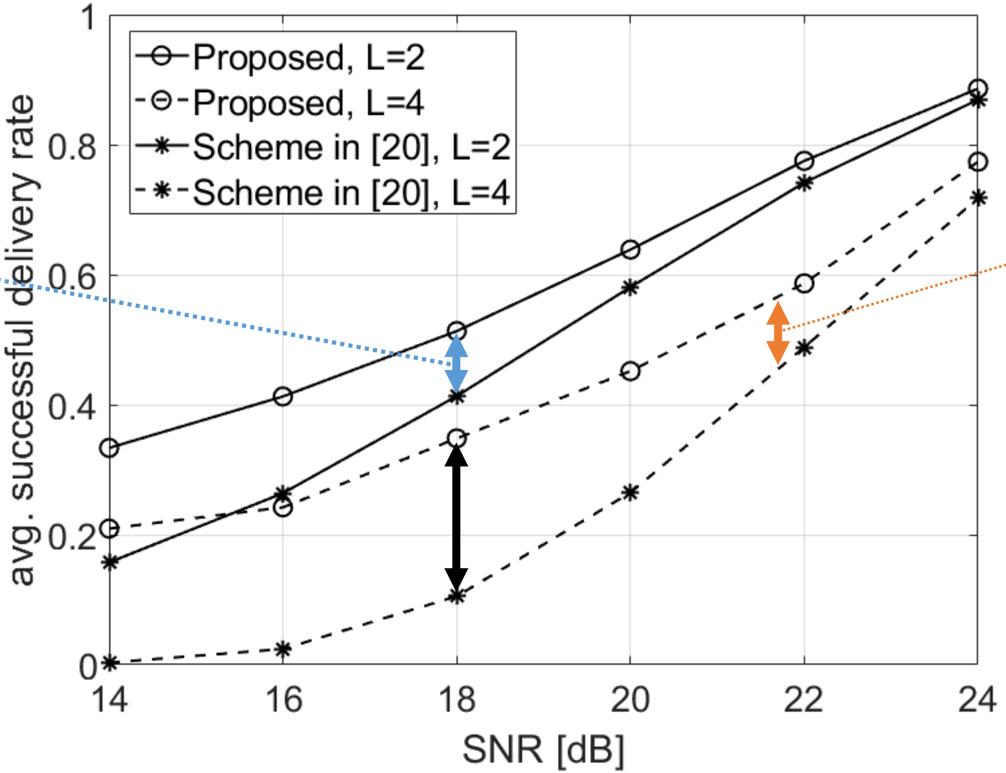
→ Their differences result from global content popularity

- As q or L or α grows, every peak becomes sharper.

→ Category-based content popularity becomes stronger than global content popularity.

Numerical Results – Avg. Successful Delivery Rates (1)

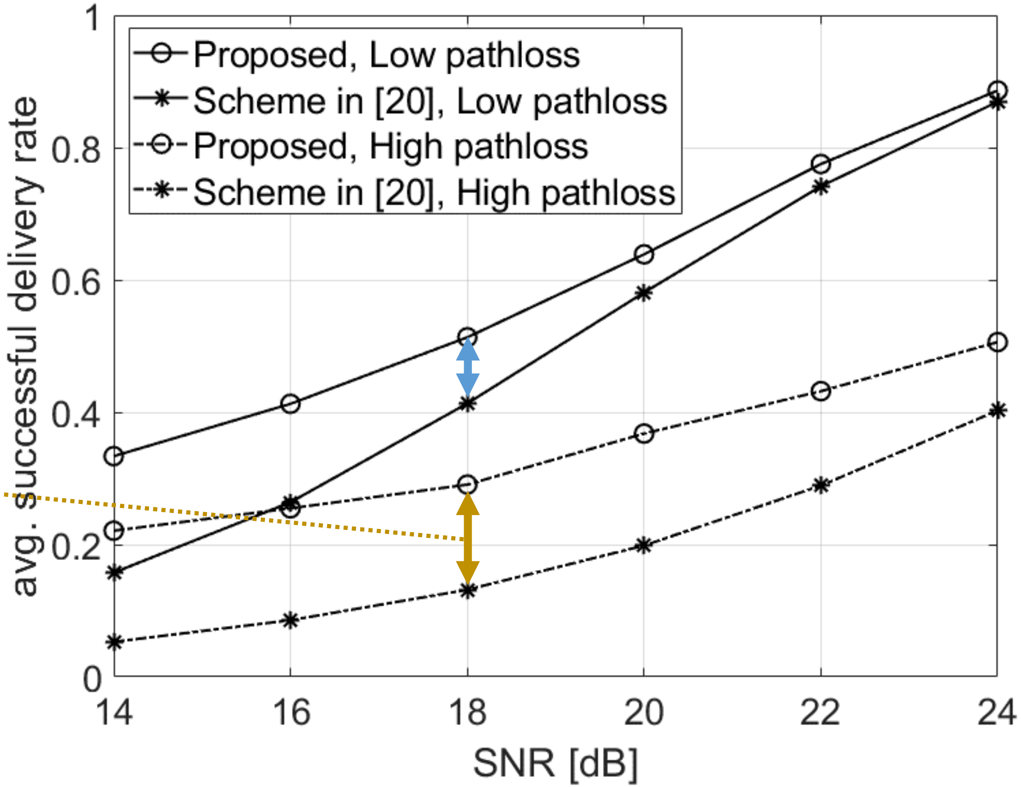
Performance gain becomes larger as L grows.



Performance gain becomes smaller as SNR grows.

➔ The proposed scheme outperforms the caching policy optimized for one-shot requests

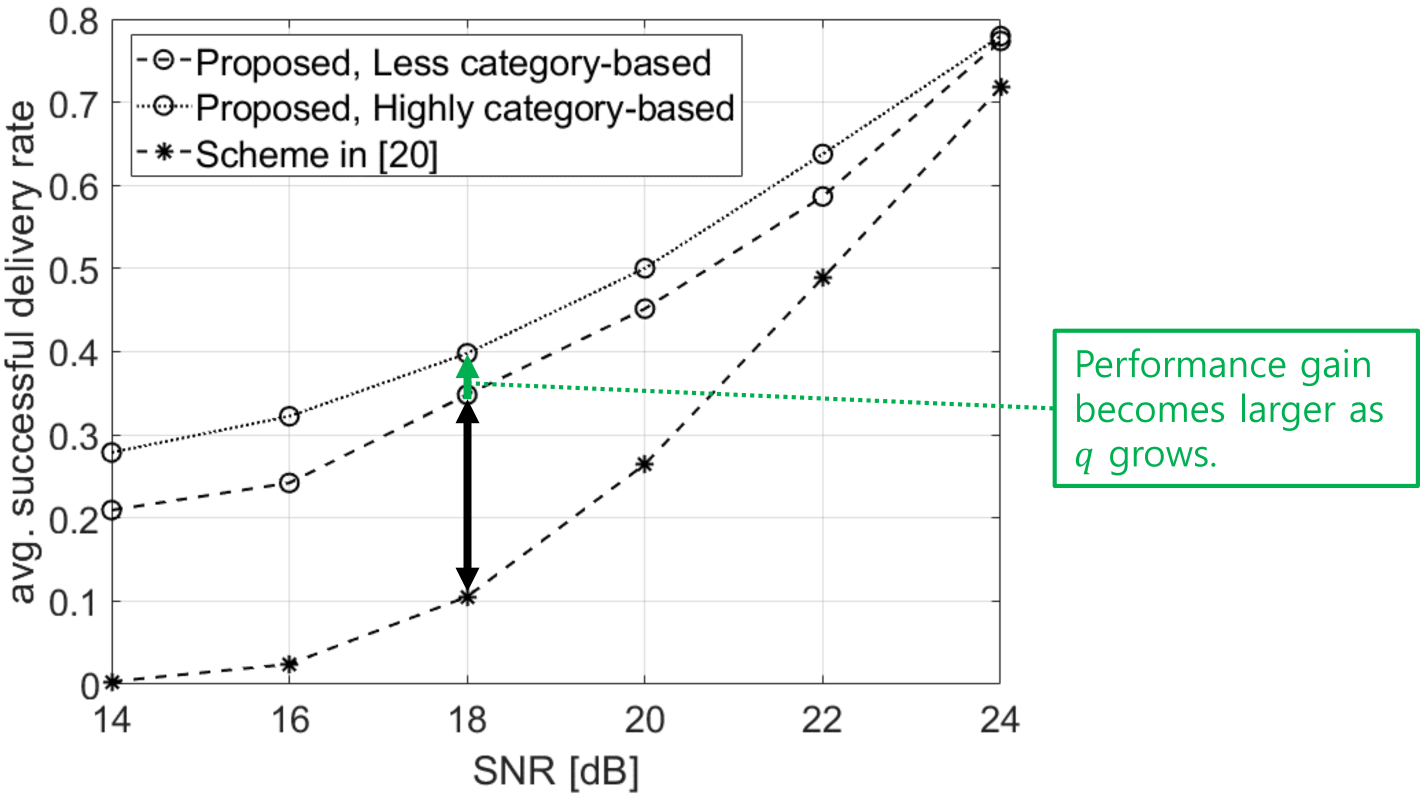
Numerical Results – Avg. Successful Delivery Rates (2)



Performance gain becomes larger as α grows.

➔ The proposed scheme outperforms the caching policy optimized for one-shot requests

Numerical Results – Avg. Successful Delivery Rates (3)



➔ The proposed scheme outperforms the caching policy optimized for one-shot requests

Conclusion

- The probabilistic caching policy that maximizes the minimum average successful delivery rates among all users is proposed when users request different numbers of categorized contents.
- The proposed scheme captures the essential characteristics of video delivery:
  - Contents in the same category have higher relevance.
  - Different users demand different content consumption.
- The iterative algorithm optimizing the dual-variable subproblems can solve the original optimization problem with the guarantee of convergence.
- The impacts of categorized contents and consecutive user demands on the caching policy are clearly shown by numerical results.