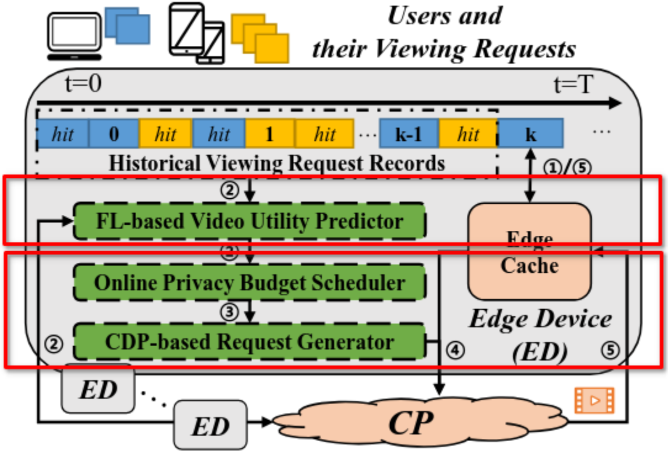
Privacy-Preserving Video Fetching

Threshold-based Online Algorithm and CDP-based Video Pre-fetching

Kartik Saini
Roll Number: 2201103
Project Guide: Dr. Radhika Sukapuram

25th February, 2025

Overview



Covered so far

To be covered

Introduction

Online video streaming is widespread, but user requests can expose private preferences to content providers (CPs).

- The Privacy-Preserving Video Fetching (PPVF) framework, as introduced in [1], uses edge devices (EDs) to pre-fetch and cache videos.
- Goal: Protect user privacy while maintaining caching efficiency.

Threshold-based Online Algorithm: Introduction

What is it?

- A mechanism to select candidate videos for pre-fetching on EDs in real-time, as proposed in [1].
- Operates online, makes decisions in real time as video requests arrive.

Goal in PPVF:

- Maximize video utility within a limited privacy budget.

Online Privacy Budget Allocation Algorithm for ED e

How it Works:

1. Start with an empty candidate set for each pre-fetching slot.
2. Randomly pick a video and compute its utility-to-cost ratio: $\frac{\lambda_{e,i}^{k}}{\epsilon_{e,i}}$.
3. Compare to a threshold $\Theta_e(\gamma_{e,i}^{k})$, where γ is the used privacy budget fraction.
4. If ratio exceeds threshold, add video to the set.
5. Threshold rises as budget is used, increasing selectivity.
6. Stop when set reaches capacity f_e or no videos qualify.

Threshold-based Online Algorithm: Achievements

- Selects high-utility videos efficiently under privacy constraints.
- **Output:** Produces the candidate set $\mathcal{A}_e^k$ of videos for pre-fetching at each time slot k.

Introduction to Correlated Differential Privacy (CDP)

What is CDP?

- Differential Privacy (DP) adds noise to protect privacy in independent datasets.
- Correlated Differential Privacy (CDP) extends DP to handle correlated data by adjusting noise based on correlations.
- This prevents privacy leaks that can occur when correlations are ignored in standard DP.

Why is it used?

- In video pre-fetching, CDP generates noisy requests that obscure user preferences while accounting for video correlations, as developed in [1].
- It ensures effective privacy protection without degrading caching performance.

Exponential Mechanism in Differential Privacy

The Exponential Mechanism is a method to select an output privately:

- Given a set of possible outputs (e.g., videos to pre-fetch).
- Each output has a utility score (e.g., video utility $\lambda_{e,i}^{k}$).
- Select output i with probability proportional to $\exp\left(\frac{\epsilon \cdot \text{utility}(i)}{2 \cdot \Delta u}\right)$, where ϵ is the privacy budget and Δu is the sensitivity.

This ensures differential privacy while favoring high-utility outputs.

CDP-based Video Pre-fetching: Introduction

What is it?

- Generates noisy pre-fetching requests from the candidate set using the Exponential Mechanism, as proposed in [1].
- Uses Correlated Differential Privacy (CDP) to adjust noise based on video relationships.

Goal in PPVF:

- Hide true user preferences from CPs while enabling effective caching.

Online Privacy-preserving Videos Pre-fetching Algorithm for ED e

How it Works:

1. Take candidate set $\mathcal{A}_e^k$ from the previous algorithm.
2. Update correlation matrix Ψ_e^k using historical request data.
3. Compute correlated sensitivity $\Delta\lambda_{e,gc}^k$ based on video relationships.
4. Use the Exponential Mechanism to select videos from $\mathcal{A}_e^k$ with probability $\propto \exp\left(\frac{\epsilon_e^k \cdot \lambda_{e,i}^k}{2 \cdot \Delta\lambda_{e,gc}^k}\right)$ up to capacity f_e.

CDP-based Video Pre-fetching: Achievements

- Preserves privacy by obscuring user preferences.
- CDP ensures noise fits video correlations, avoiding excess distortion.
- Maintains caching efficiency.
- **Output:** Produces the pre-fetching decision vector $\mathbf{x}_e^k$ for selecting videos to pre-fetch.

Conclusion

The PPVF framework, as introduced in [1], balances privacy and utility using:

- Threshold-based Online Algorithm: Efficient video selection for caching at the edge device.
- CDP-based Pre-fetching: Privacy-preserving requests with correlation-aware noise via the Exponential Mechanism.

Together, these algorithms create a privacy-preserving edge caching system that effectively balances user privacy and service quality.

Further Work

- Scale down the code to understand the implementation details.
- Modify the implementations to accommodate service caching.
- Verify the claimed results through simulations or experiments.

References I

Xianzhi Zhang, Yipeng Zhou, Di Wu, Quan Z. Sheng, Miao Hu, and Linchang Xiao.
Ppvf: An efficient privacy-preserving online video fetching framework with correlated differential privacy.
arXiv preprint arXiv:2408.14735, 2024.

Acknowledgement

Thank you!

Understanding the Threshold Function

Purpose: Determines whether a video is selected for the candidate set based on its utility-to-cost ratio. **Definition:**

$$\Theta_e(\gamma) = \begin{cases} L_e & \text{if } 0 \le \gamma \le \Gamma_e \\ \frac{L_e}{\exp(1)} \left(\frac{U_e \cdot \exp(1)}{L_e} \right)^{\gamma} & \text{if } \Gamma_e < \gamma \le 1 \end{cases}$$

Key Variables:

- γ: Fraction of privacy budget used for a video.
- $\Gamma_e = \frac{1}{1+\ln\left(\frac{U_e}{L_e}\right)}$: Point where threshold starts increasing.
- L_e, U_e: Minimum and maximum utility-to-cost ratios.

Behavior:

- For $\gamma \le \Gamma_e$, threshold is constant at L_e.
- For $\gamma > \Gamma_e$, threshold increases exponentially towards U_e.

Intuition:

- Early on (low γ), select videos with ratio above L_e.
- As budget depletes (high γ), require higher ratios.
- Prioritizes high-utility videos when budget is limited.

Correlation Matrix in CDP: Introduction

What is it?

- $\Psi_e^k = [\Psi_{e,i,j}^k]^{I \times I}$: Matrix of Pearson correlation coefficients between videos i and j at time k for edge device e, where I is the total number of videos.
- $\Psi_{e,i,j}^k$: Measures the linear relationship between the utility sequences $\{\lambda_{e,i}^1, \ldots, \lambda_{e,i}^k\}$ and $\{\lambda_{e,j}^1, \ldots, \lambda_{e,j}^k\}$, where $\lambda_{e,i}^k$ is the predicted utility of video i at time k.

Purpose in CDP:

- Captures relationships between videos based on user request patterns [1].
- Adjusts noise in Correlated Differential Privacy (CDP) to reflect these correlations, enhancing privacy without excessive utility loss.

Overview of Calculation:

- Uses historical utility data to compute correlations incrementally.
- Involves updating cumulative sums and applying the Pearson correlation formula.

Correlation Matrix in CDP: Calculation Details

Detailed Calculation:

- **Historical Sums (initialized at zero for $k = 0$):**
  - $\alpha_{e,i}^{k} = \sum_{m=1}^{k} \lambda_{e,i}^{m}$: Cumulative sum of utilities for video i.
  - $\sigma_{e,i}^{k} = \sum_{m=1}^{k} (\lambda_{e,i}^{m})^2$: Cumulative sum of squared utilities for video i.
  - $\psi_{e,i,j}^{k} = \sum_{m=1}^{k} \lambda_{e,i}^{m} \lambda_{e,j}^{m}$: Cumulative sum of utility products for videos i and j.
- **Update Rules:**
  - $\alpha_{e,i}^{k} = \alpha_{e,i}^{k-1} + \lambda_{e,i}^{k}$
  - $\sigma_{e,i}^{k} = \sigma_{e,i}^{k-1} + (\lambda_{e,i}^{k})^2$
  - $\psi_{e,i,j}^{k} = \psi_{e,i,j}^{k-1} + \lambda_{e,i}^{k} \lambda_{e,j}^{k}$
- **Correlation Coefficient:**

$$\Psi_{e,i,j}^{k} = \frac{k \cdot \psi_{e,i,j}^{k} - \alpha_{e,i}^{k} \cdot \alpha_{e,j}^{k}}{\sqrt{k \cdot \sigma_{e,i}^{k} - (\alpha_{e,i}^{k})^2} \cdot \sqrt{k \cdot \sigma_{e,j}^{k} - (\alpha_{e,j}^{k})^2}}$$

- This matches the Pearson correlation formula for the utility sequences up to time k.

Correlation Matrix in CDP: Intuition

Intuition:

- $\Psi^k_{e,i,j} \approx 1$: Strong positive correlation; similar utility trends.
- $\Psi^k_{e,i,j} \approx -1$: Strong negative correlation; opposite trends.
- $\Psi^k_{e,i,j} \approx 0$: No linear correlation.
- Guides noise adjustment in CDP for effective privacy [1].

Sensitivity in CDP

What is it?

- Measures how much output (utility) changes with input data changes.
- In CDP, accounts for video correlations.

Correlated Video Sensitivity:

$$\Delta\lambda_{e,i}^{k} = \sum_{j \in \mathcal{A}_e^k} \left(\Psi_{e,i,j}^{k} \cdot \left\| h_e(i, t^k | \mathcal{V}_e^k, \boldsymbol{\theta}) - h_e(i, t^k | \mathcal{V}_{e,-j}^k, \boldsymbol{\theta}) \right\|_1 \right)$$

- Impact on video i's utility when removing video j's requests.

Global Sensitivity:

$$\Delta\lambda_{e,gc}^{k} = \max_{i \in \mathcal{A}_e^k} \Delta\lambda_{e,i}^{k}$$

- Maximum sensitivity across all candidate videos.

Why it Matters:

- Sets noise level in the Exponential Mechanism.
- Higher sensitivity needs more noise for privacy.
- CDP calibrates noise to correlations, optimizing utility.