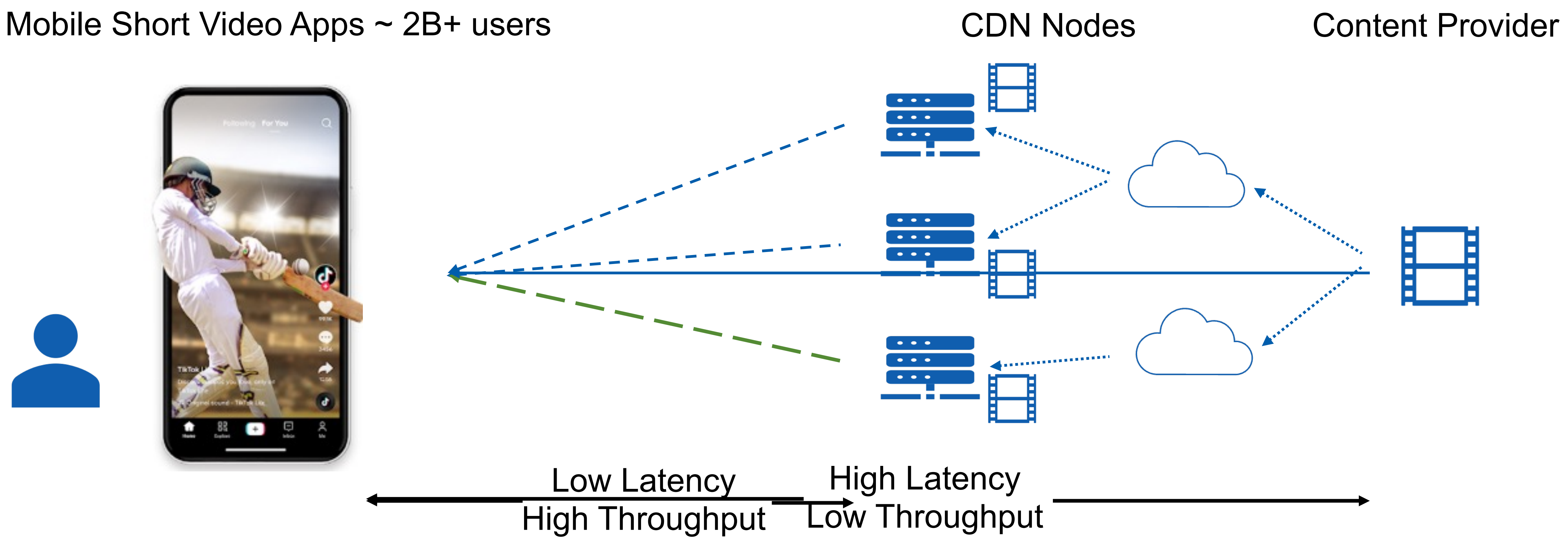
Bridge the Gap Between QoS and QoE in Mobile Short Video Service: a CDN Perspective

Chuanqing Lin, Yangguang Liang, Fuhua Zeng, Zhipeng Huang, Xiaodong Li, Jingyu Yang, Yu Tian, Gerui Lv, Qinghua Wu, Zhenyu Li, and Gaogang Xie

Institute of Computing Technology, Chinese Academy of Sciences (ICT, CAS)
University of Chinese Academy of Sciences (UCAS)
Purple Mountain Laboratories
Computer Network Information Center, Chinese Academy of Sciences (CNIC, CAS)

Background: Content Delivery Network (CDN)

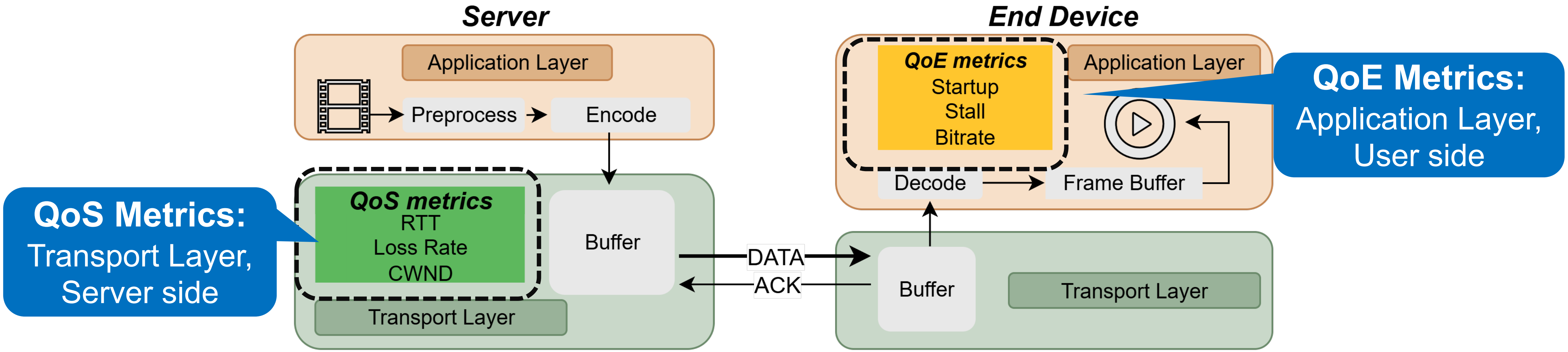
- CDNs support emerging mobile short video services.



CDN's goal is to optimize **performance** for serving applications by **scheduling**.

Performance Metrics: QoE ≠ QoS

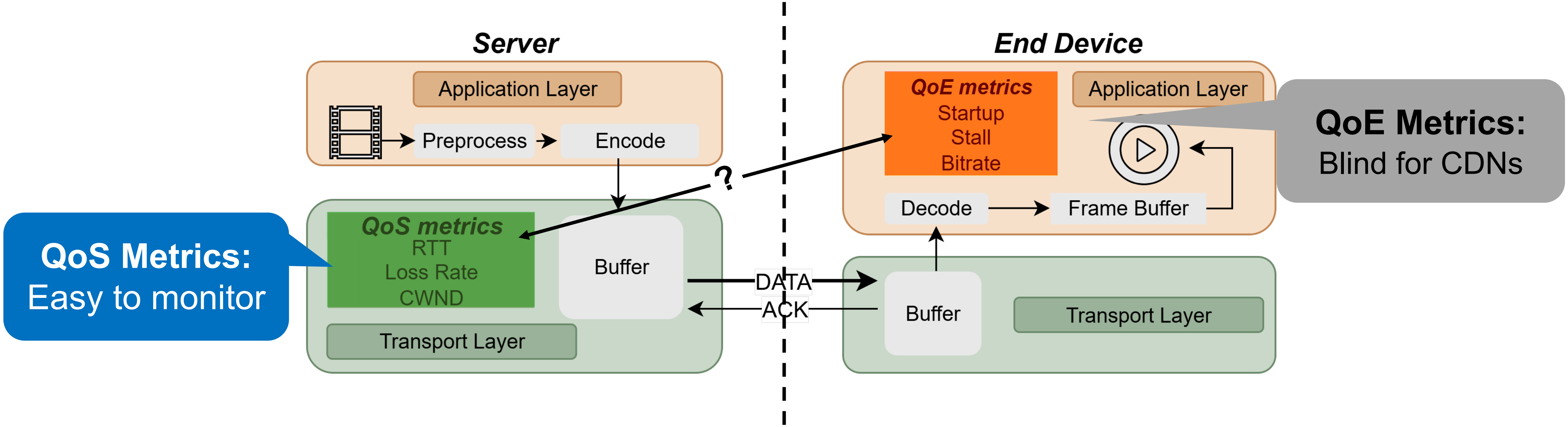
- Users care **Quality of Experience** (QoE)
  - Startup delay, stall, bitrate selection, etc. in playback
- CDNs monitor **Quality of Service** (QoS)
  - RTT, loss rate, throughput, etc. in network stack



Core Problem: QoS and QoE are mismatched.

Motivation: Bridge the QoS-QoE Gap

- CDN cannot observe QoE directly.
- CDN's efforts to improve QoE cannot always take effects.
  - $QoS \uparrow \nRightarrow \mathrm{QoE} \uparrow$



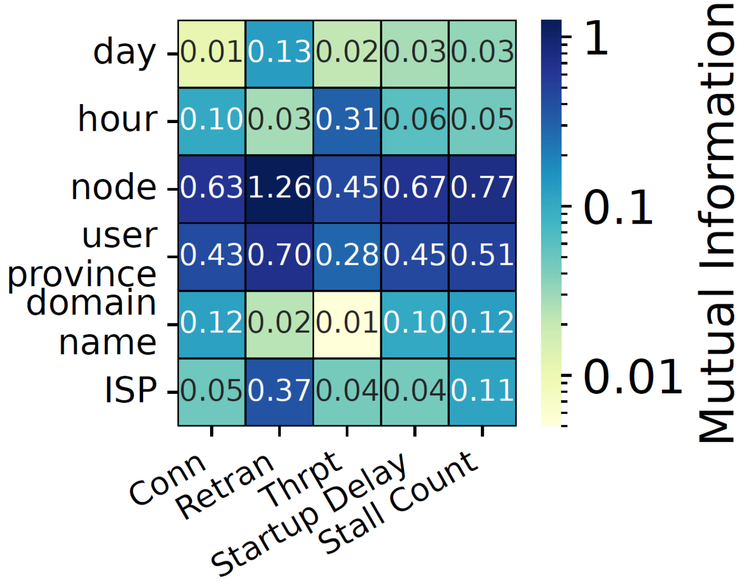
Goal: Construct QoS-QoE mappings for **scheduling** to achieve better QoE

Complex QoS-QoE Relationship

- Q: Can we simply use a **supervised model** to predict QoE from QoS?
- A: Data-driven prediction model requires complex scenario features as input.
  - Scenario features have strong correlation with QoE.

Features	Startup Delay / %	Stall Count / %
QoS	6.0	23.7
QoS + Scenarios	3.1 1.9×	8.7 2.7×

XGBoost model's prediction accuracy with different input features.



Mutual Information between metrics and scenario features.

Design challenge: Complex scenario features reduce interpretability.

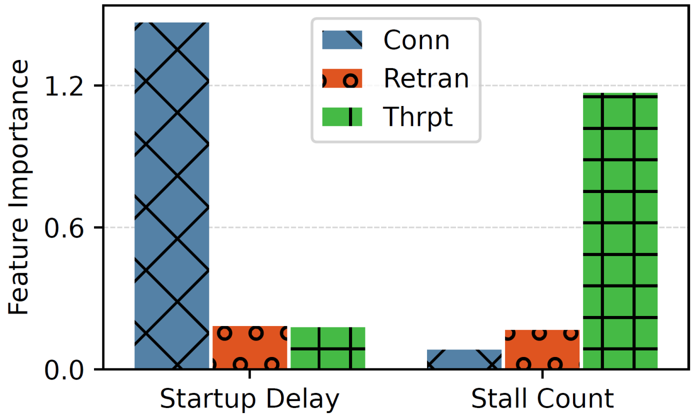
Core Idea: From modeling to clustering

PATTERN	GROUP	APPLY
Visualize QoS-QoE mappings Use Quasi-Experimental Design (QED) to generate a large set of QoS-QoE mapping patterns.	**Cluster** to summarize scenarios Apply K-Medoids clustering to group patterns and extract representative "meta-scenarios".	**Schedule** with predicted QoE Build cluster-specific prediction models to power a new QoE-aware CDN scheduling objective.

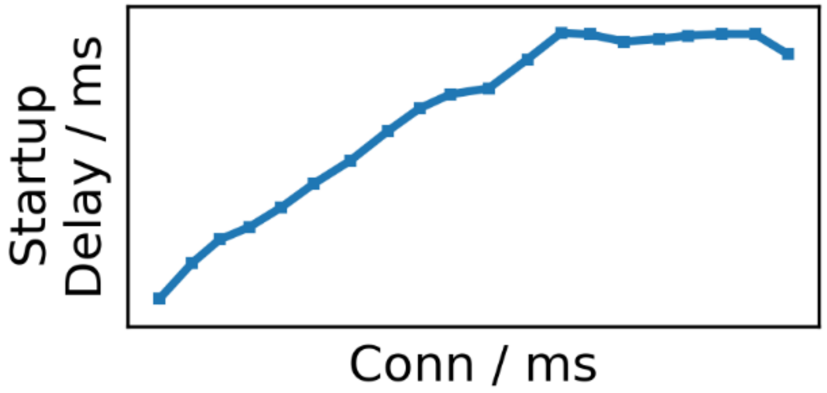
Step 1: Uncovering Individual Patterns via QED

- Observation: QoE metrics are strongly related to **one** corresponding QoS metrics.
- Quasi Experimental Design (QED): Construct univariate functions under different scenarios.

$$QoE_i = f(QoS_i | scenarios)$$



Feature importance for different QoE metrics.



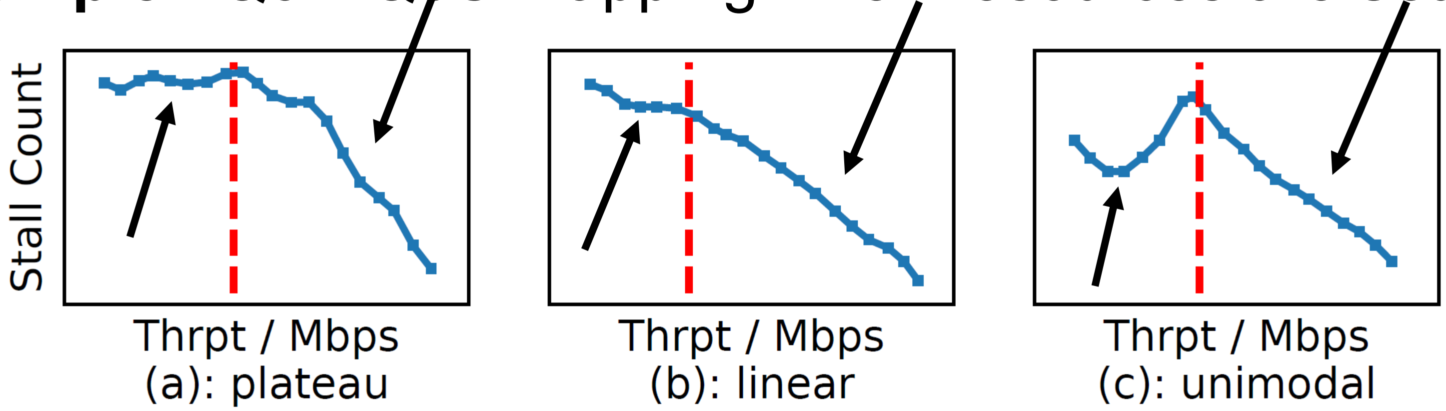
Example for the pattern of univariate function.

Key result: 1,800+ visualized patterns under corresponding scenarios.

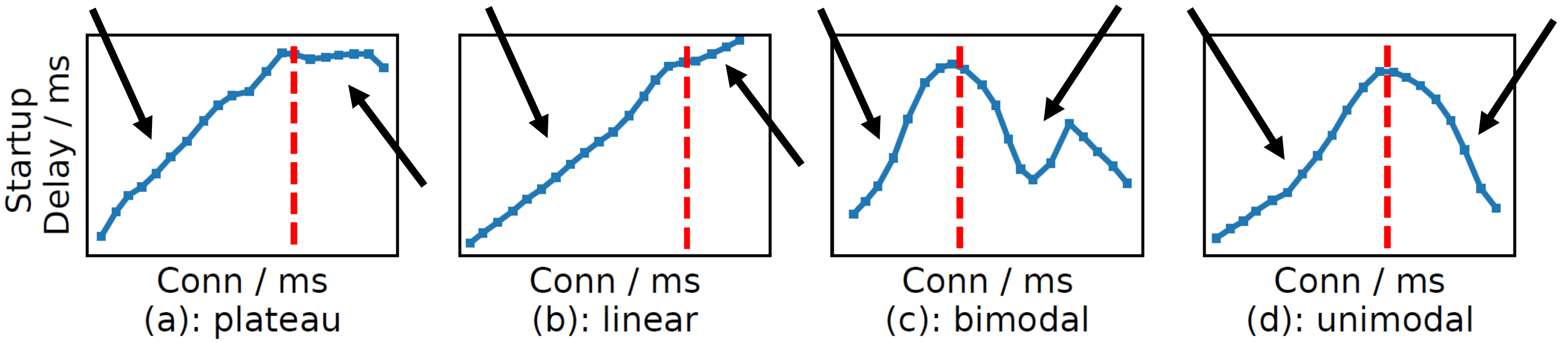
Step 2: Summarizing Scenarios with Clustering

- K-Medoids **clustering** with Dynamic Time Warping (DTW) distance.
- **Observation**:
    1. Approximately **linear** QoE-QoS mapping when resources are **abundant**.
    2. **Non-linear and complex** QoE-QoS mapping when resources are **scarce**.

<throughput, stall count>



<connection setup time, video startup delay>



Implication for Scheduling

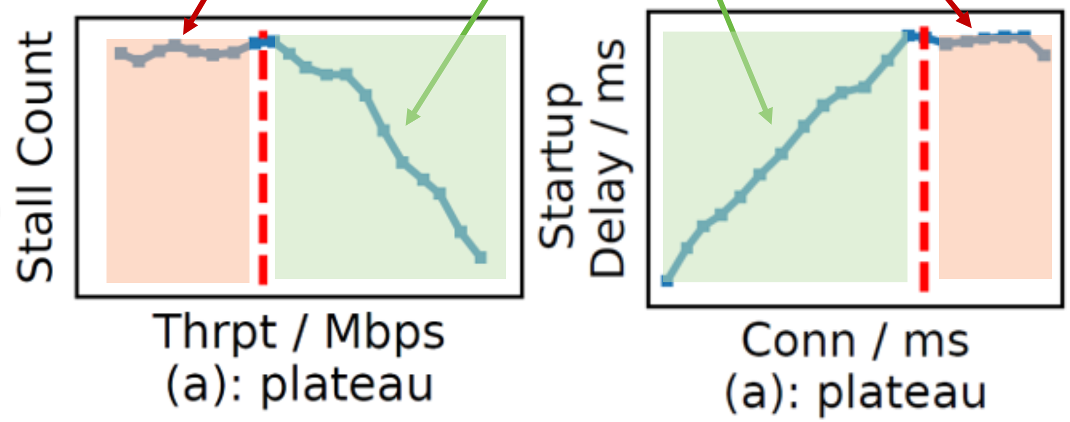
- Insight 1: **Performance** differentiated node utilization strategies.
  - Maximize utilizing efficiency for *high-performance nodes*.
- Insight 2: **Time** differentiated scheduling strategies.
  - Maximize QoE for *users*.

Low QoS, conservative strategies, less QoE loss

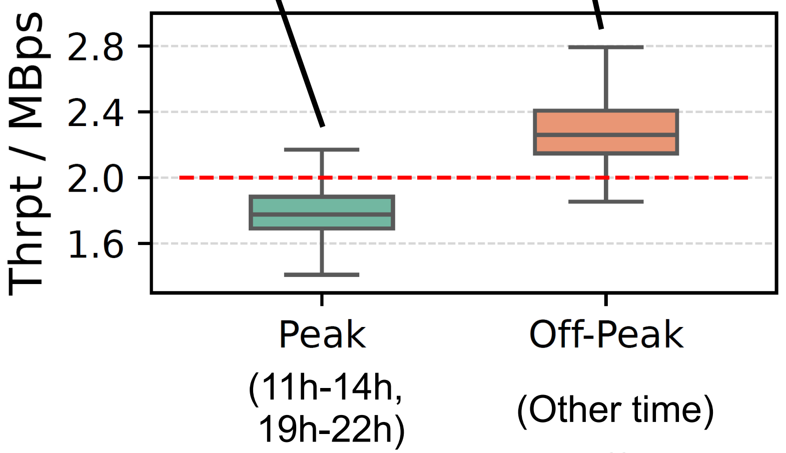
QoE Almost No Change → Poorer QoS is ok.

High QoS, aggressive strategies, better QoE gain.

QoS Improvement Zone → Better QoS helps.



Plateau shaped QoS-QoE mapping pattern.



Throughput distribution at different times.

Step 3: QoE-Aware CDN Scheduling

- CDN scheduling is an optimization problem to find the scheduling strategy X.

$$\min_X \boxed{D_{ij}} \cdot X_{ij}$$

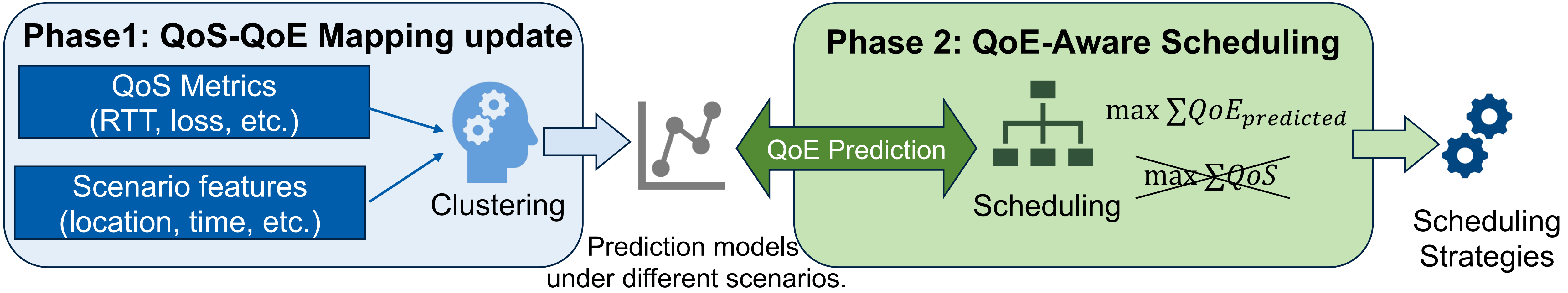
Cost function D_{ij} is predicted QoE, instead of QoS.

$$s.t. \sum_j X_{ij} \le C_i, \forall i \in N;$$

CDN node capacity constraints.

$$\sum_i X_{ij} = T_j, \forall j \in M.$$

Traffic allocation constraints.



Evaluation: Clustering-based Prediction

- Real-world collected Dataset:
  - QoE: startup delay, stall count
  - QoS: RTT, loss rate, throughput

- Baseline: **XGBoost**
- Prediction error: **5.5%** ▼
  - Clustering-based method is more **interpretable**.

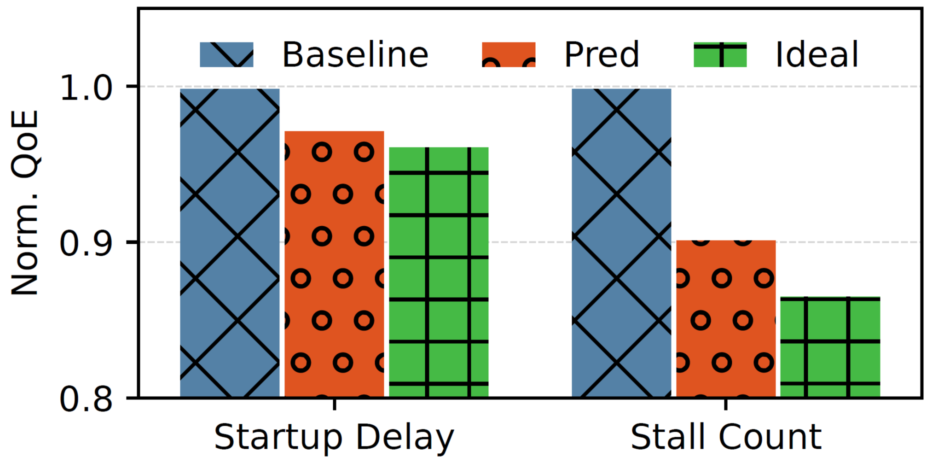
Cluster	1	2	3	4	5	6	7	8	AVG	XGBoost
Startup delay / %	3.2	2.3	2.3	2.8	3.2	2.5	4.4	2.5	**2.9**	**3.1**
Stall count / %	7.7	10.1	9.8	6.9	7.0	8.0	-	-	**8.3**	**8.7**

Comparison of Mean Absolute Prediction Error (MAPE)
for each clustering group, overall and XGBoost model.

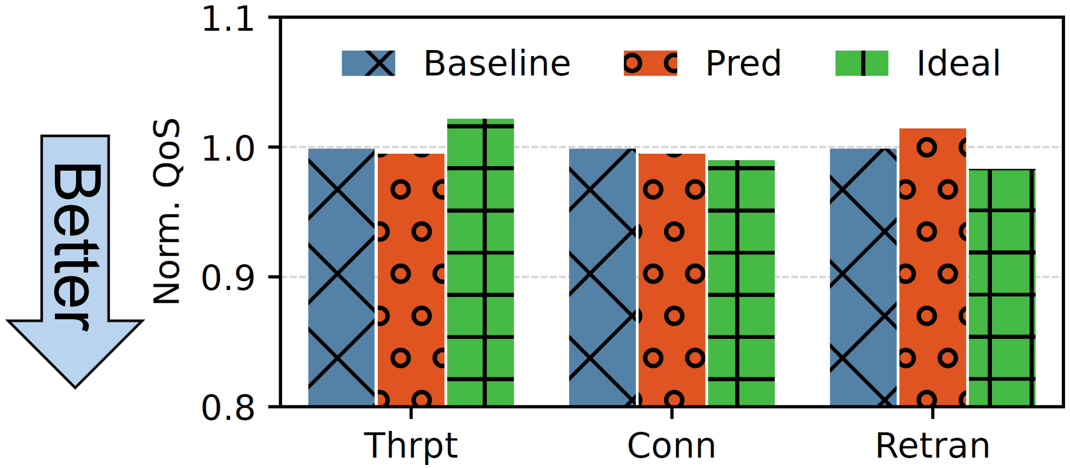
Evaluation: QoE-aware Scheduling

- Testbed: a simulation environment for CDN traffic scheduling
- Baseline: QoS-based scheduling
- Ideal: accurate QoE-aware scheduling (Optimal)

- QoE-aware scheduling achieves **9.9%** QoE ▲
  - under comparable QoS supplement.



Normalized QoE punishment.



Normalized QoS punishment.

Summary

- **Observation:** QoE and QoS are mismatched, and relationships are complex.
  - The mapping patterns are varied among different scenarios.

- **Solution:** a clustering-based prediction framework.
  - Accurately predicting the QoE performance from QoS metrics under different scenarios with interpretability.

- **Impact:** QoE-aware CDN scheduling
  - Providing additional QoE improvement under comparable QoS supplement.

中国科学院计算技术研究所 INSTITUTE OF COMPUTING TECHNOLOGY, CHINESE ACADEMY OF SCIENCES
中国科学院大学 University of Chinese Academy of Sciences
紫金山实验室 Purple Mountain Laboratories
中国科学院计算机网络信息中心 Computer Network Information Center, Chinese Academy of Sciences

Thanks!

Q & A

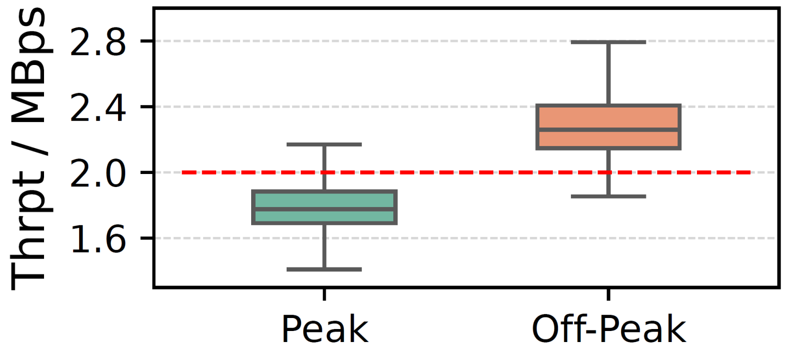
For any further questions, please contact:

Chuanqing Lin (ICT, CAS)

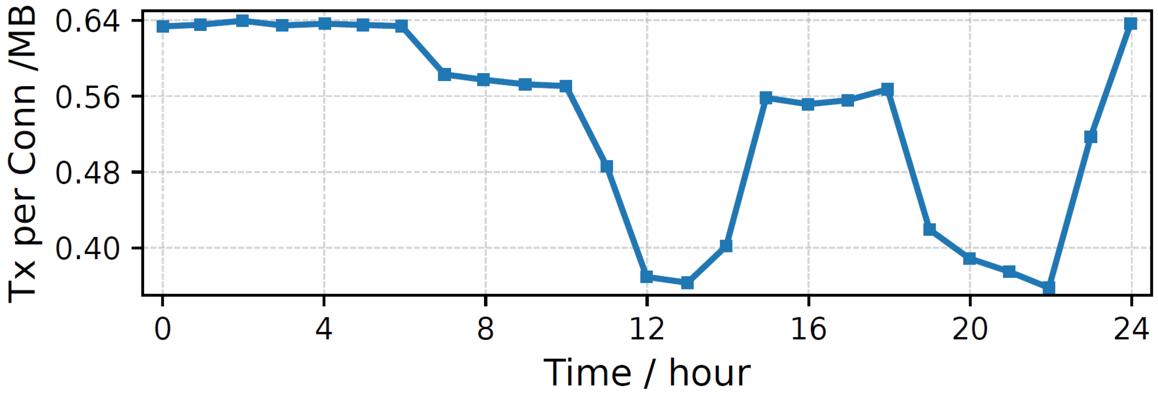
linchuanqing20b@ict.ac.cn

Appendix：
Resource Scarcity is the Key Differentiator

- Transition points widely exist.
  - $Thrpt_{tran} = \frac{Q_3(Thrpt_{peak})+Q_1(Thrpt_{off-peak})}{2}$
- Potential factors: user-side bit rate adaption strategies.



Throughput division in traffic peak (11h-14h, 19h-22h) and off-peak.



Mean amount of data transmission per TCP stream within a day.