SFU SIMON FRASER UNIVERSITY
ENGAGING THE WORLD

ITSVA: Toward 6G-Enabled Vision Analytics over Integrated Terrestrial-Satellite Network

Miao Zhang, Jiaxing Li, Jianxin Shi, Yifei Zhu, Lei Zhang, Hengzhi Wang

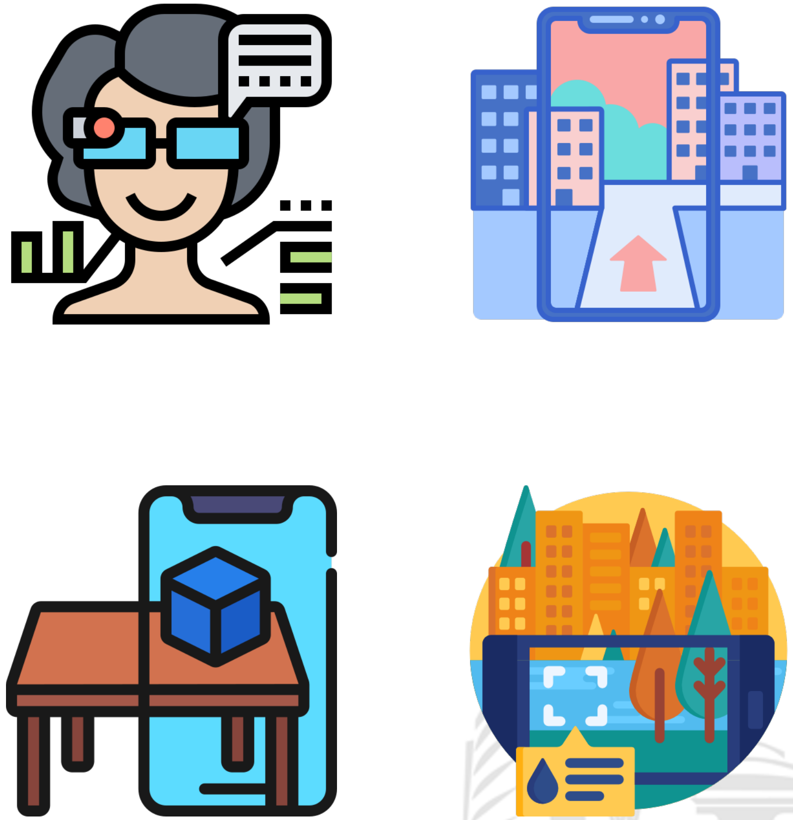
BACKGROUND

❑ Mobile Vision Analytics

Mobile vision analytics (MVA) enables machines to understand the physical world by analyzing videos captured by mobile devices in real time.

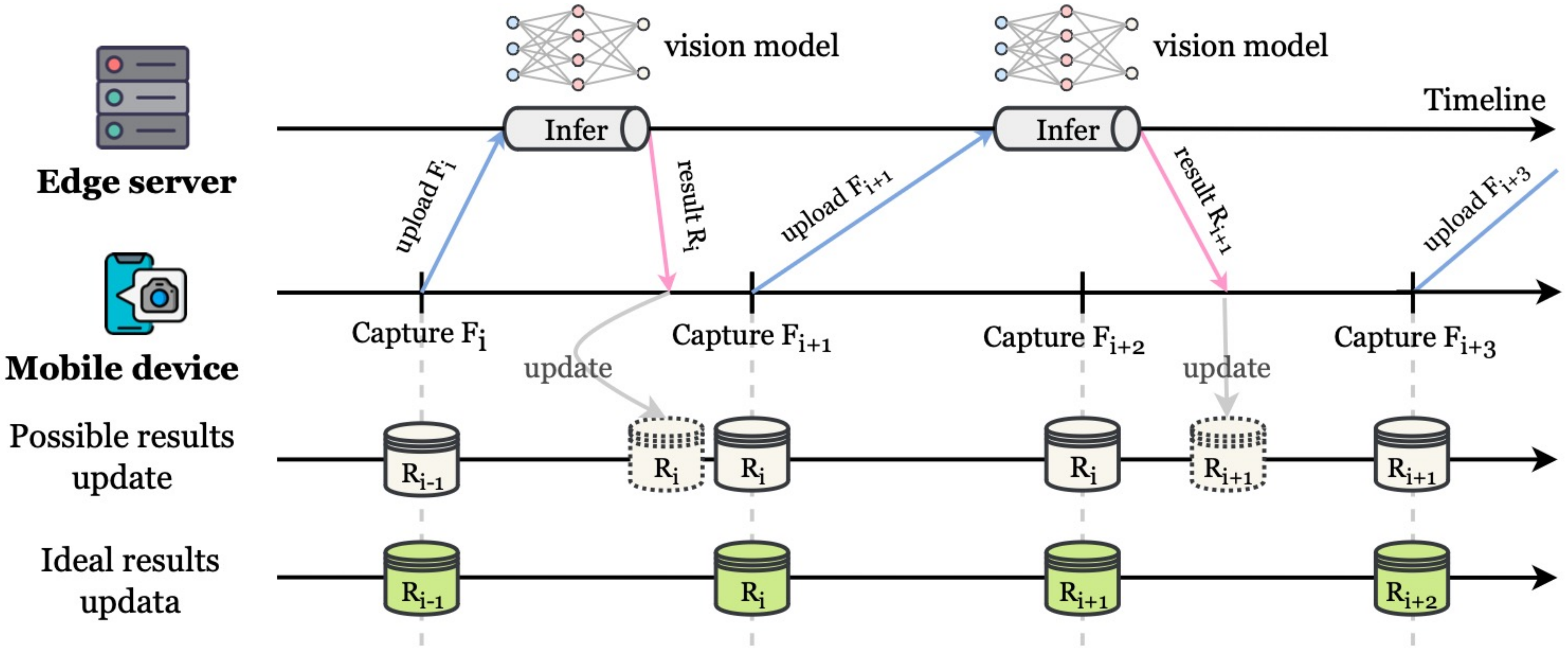
With deep neural network (DNN) based vision models, MVA can help bridge the gap between the physical and virtual worlds.

Given the constrained computational resources and heat dissipation issues of mobile devices, existing MVA systems tend to offload heavy DNN inference workloads to edge servers.



BACKGROUND

❑ Edge-Assisted Mobile Vision Analytics



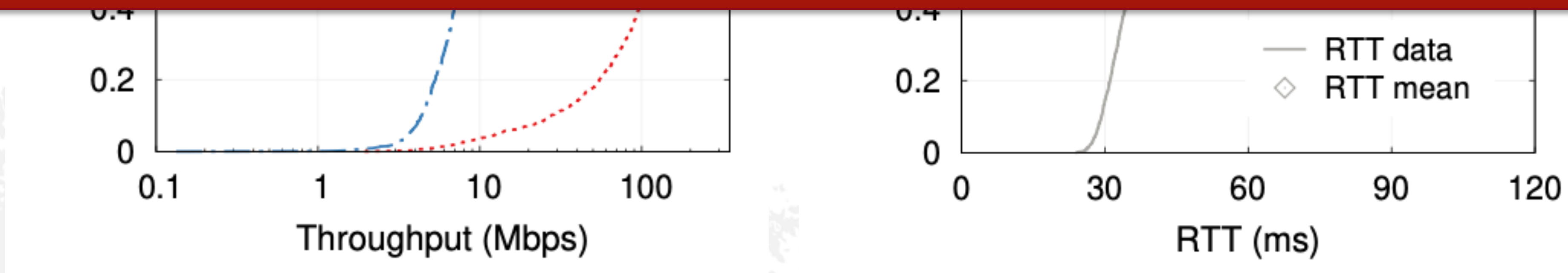
Frame Offloading Workflow of Edge-Assisted MVA

MOTIVATION

❑ Characteristics of Integrated Terrestrial and LEO Satellite Network (ITLSN)

- ❖ LEO Internet service provider: Starlink
- ❖ Mobile device: Raspberry Pi 4 Model B
- ❖ Edge server: rented from the nearest AWS local region to the mobile device
- ❖ TCP throughput measurement tool: IPerf3 utility
- ❖ Round-trip time (RTT) measurement tool: Ping utility

Today's ITLSN still cannot support high-frame-rate offloading, and specialized designs are required toward 6G-enabled MVA.

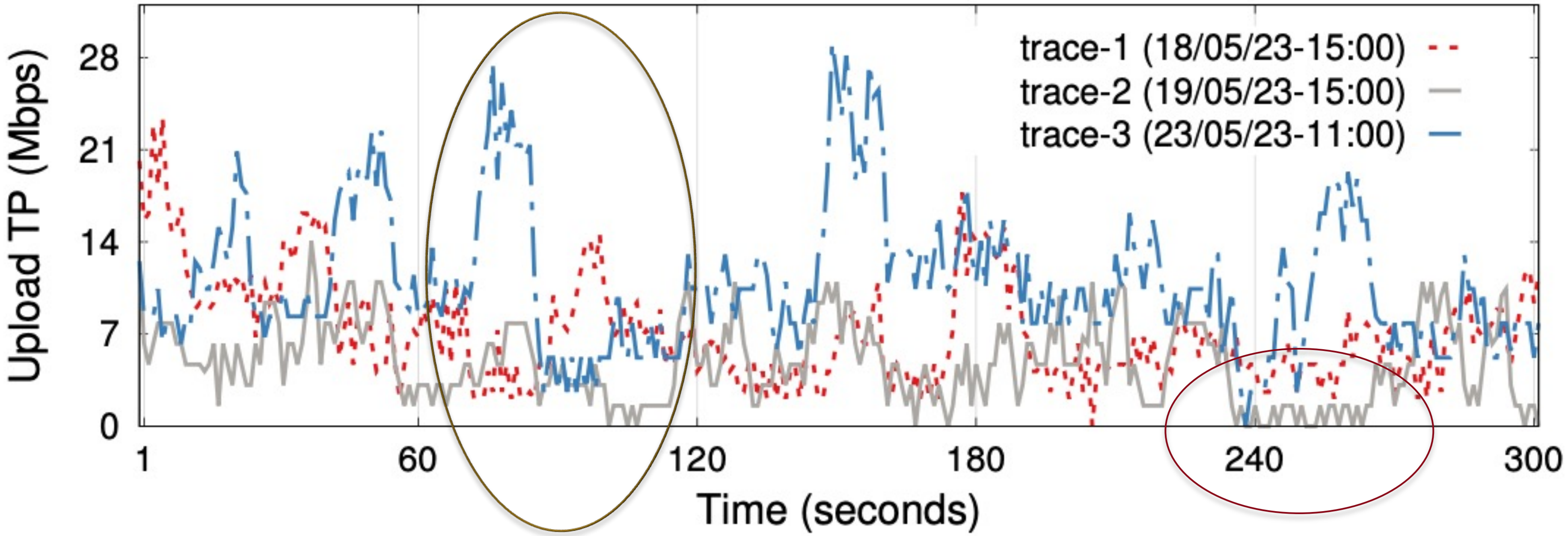


TCP throughput statistics **Network latency statistics**

MOTIVATION

❑ Characteristics of Integrated Terrestrial and LEO Satellite Network (ITLSN)

ITLSN experiences wild fluctuations in upload throughput. This calls for network-aware designs for 6G-enabled MVA to deliver consistent QoE.



Upload throughput variations over time

BACKGROUND

❑ Edge-Assisted Mobile Vision Analytics with Today’s ITLSN

How to address the network resources challenges of today’s ITLSN to achieve the high-accuracy and low-latency performance goals of edge-assisted MVA?

Solution 1: Reduce offloaded frame quality, e.g., by reducing resolution or increasing quantization parameter.

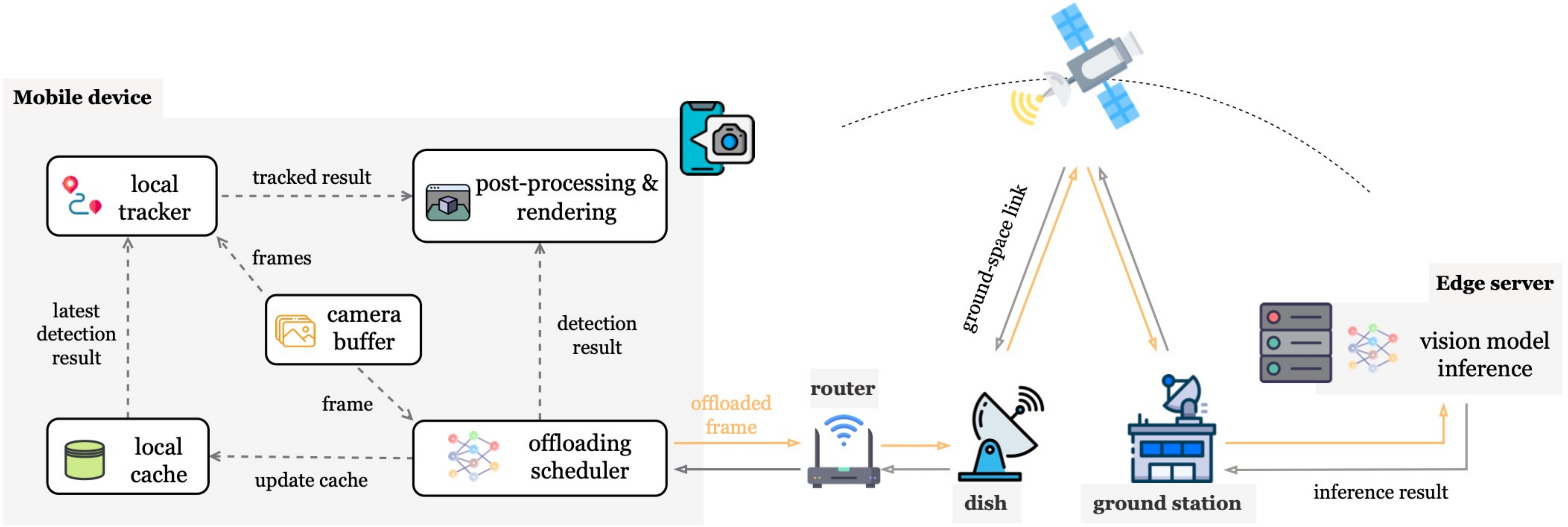
Issues: Still cannot satisfy the stringent per-frame response delay if the network has a high latency, e.g., as in ITLSN. Also, the server-side inference accuracy will be reduced due to the degraded image quality.

Solution 2: Periodically or selectively offload representative frames.

Issues: How to compensate the accuracy of unoffloaded frames and how to decide which frames to be offloaded?

SYSTEM DESIGN

❑ Overview of ITSVA



SYSTEM DESIGN

❑ Optical Flow-Based Local Tracker

Motivation: Selective frame offloading is necessary given the scarce and volatile uplink resources of today’s ITLSN.

Problem: How to compensate the analytics accuracy of unoffloaded frames?

Strawman Solution: Reuse the inference result of the latest offloaded frame?

Our Solution: Integrate an optical flow-based local tracker into ITSVA. The algorithm is lightweight and can quantify both objects and camera motion.

SYSTEM DESIGN

❑ DRL-Based Offloading Scheduler

Optimization Goals: The offloading scheduler then tunes the frame offloading interval l_t to maximize the overall accuracy while minimizing the offloaded data amount over the network.

Challenges: (1) Both too large or too small l_t can lower accuracy. (2) Video content dynamics can influence the decision. (3) The choice of l_t can have cascading influences on that of the subsequent seconds.

Deep Reinforcement Learning (DRL)-Based Solution:

$$s_t = (\vec{n}_t, \vec{u}_t, \vec{d}_{t-1}, \vec{h}_{t-1}, \delta_t, \vec{p}_t)$$

$\vec{n}_t$: Historical upload throughput; $\vec{u}_t$: Mean per−frame offloading delay; $\vec{d}_{t-1}$: Offloading decision vector of the last second; $\vec{h}_{t-1}$: offloading success vector of the last second; δ_t: the freshness of locally cached result; $\vec{p}_t$: most recent content dynamics.

SYSTEM DESIGN

❑ DRL-Based Offloading Scheduler

Methodology: For each input state s_t, the DRL agent selects an action a_t based on a trained policy π_θ $(s_t, a_t) \rightarrow [0, 1]$, where θ is the policy parameter.

For our problem, the action a_t corresponds to the offloading interval l_t .

The policy π_θ (s_t, a_t) is represented by a neural network.

The immediate reward function is defined as follows:

$$r_t = \alpha_1 A_t - \alpha_2 O_t - \alpha_3 U_t$$

Unoffloaded rate

Overall analytics accuracy for all frames captured in second t.

Actual offloaded data size

EVALUATION

❑ Evaluation Setup

We design and implement a trace-driven simulator to evaluate the performance of ITSVA.

Network traces: a large-scale network dataset incorporating 1,200 ITLSN upload traces. Each trace has a length of 60 seconds with a granularity of 1 second.

Video dataset: High-quality videos from MOTS dataset (resolution: 1920 x 1080; frame rate: 30 FPS).

Vision task: We focus on the object detection vision task and use a pre-trained YOLOv7-w6 for inference.

❑ Baselines and Metrics:

Best-effort: Offload frames back to back, unaware of the network conditions

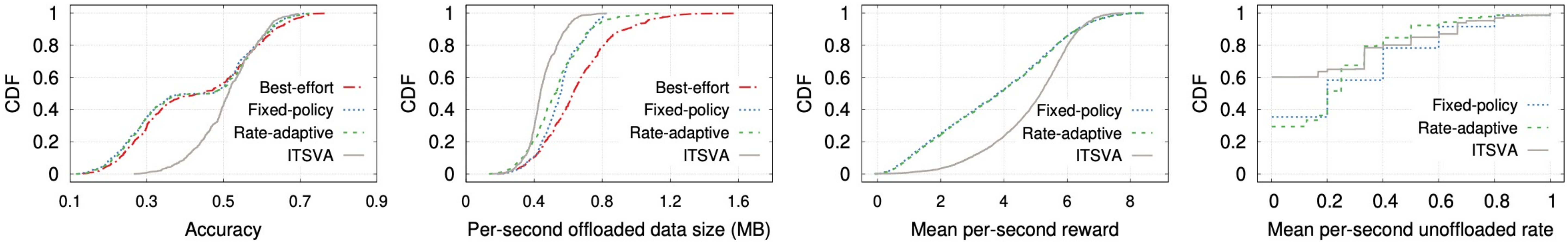
Fixed-policy: Offload frames at a fixed interval for all seconds

Rate-adaptive: Dynamically adjust the offloading interval based on the most recent network observations

Primary valuation metrics: accuracy and per-second offloaded data size.

EVALUATION

❑ Evaluation Results

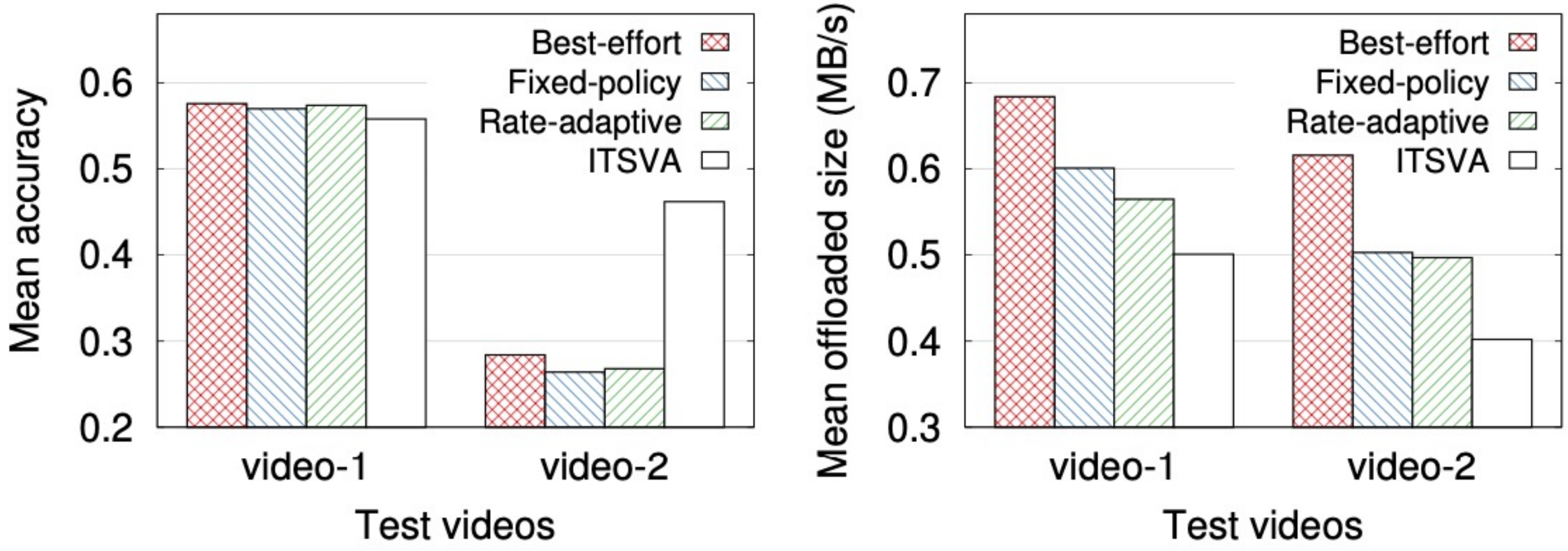


Performance comparison of different solutions (statistics of all video-trace pairs in the test set)

ITSVA attains the highest overall accuracy with significantly reduced network data transfer overhead.

EVALUATION

- Evaluation Results



Performance comparison on different test videos

SUMMARY

- Today’s ITLSN still cannot support high-frame-rate offloading, and specialized designs are required towards 6G-enabled MVA.
- ITLSN experiences wild fluctuations in upload throughput. This calls for network-aware designs for 6G-enabled MVA to deliver consistent QoE.
- By combining knowledge extracted from current network conditions, video content dynamics, and local cache status, ITSVA is able to make rewarding offloading decisions.

THANK YOU

Q & A