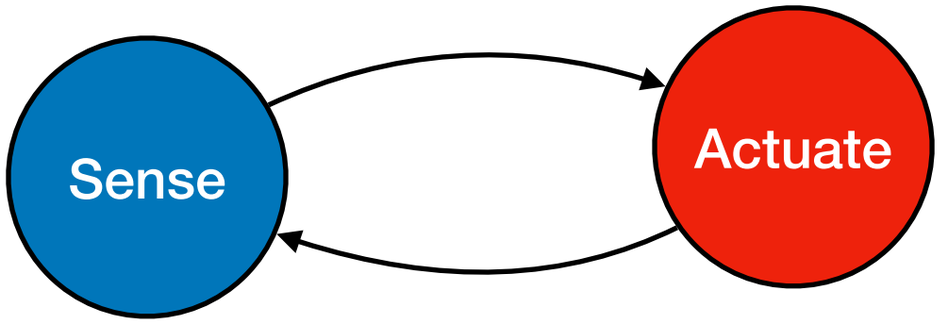
DIDL 2021

RAMPS: Next Generation Platform for Real Time and Resilient IoT Analytics using MmWave and Programmable Switches

Vishal Shrivastav, Dimitrios Koutsonikolas, Saurabh Bagchi

Autonomous Internet-of-Things (IoT)

- Autonomous IoT devices are ubiquitous



IoT Analytics

- IoT devices capture large amounts of data every second
- …need computing resources to **analyze** and generate responses

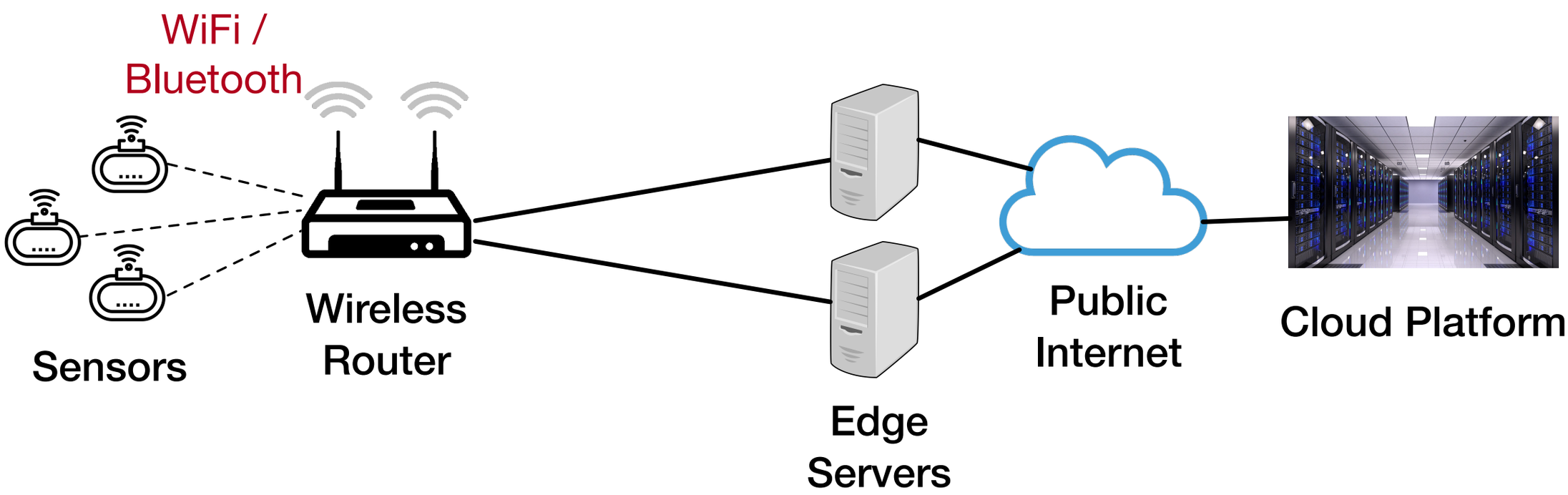
IoT Analytics

- IoT devices capture large amounts of data every second
- …need computing resources to **analyze** and generate responses
- But, IoT devices typically are computation and power constrained

IoT Analytics

- IoT devices capture large amounts of data every second
- …need computing resources to **analyze** and generate responses
- But, IoT devices typically are computation and power constrained
- State-of-the-Art: **A Distributed IoT Analytics Platform**

Distributed IoT Analytics



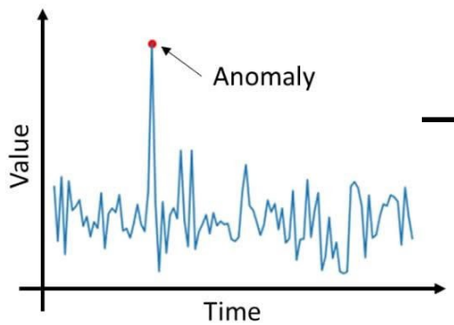
State-of-the-art Distributed IoT Analytics Platform

Our Motivation: Real Time IoT Analytics

- Several IoT applications need to analyze and generate responses in **real time**

Raise an alarm

Analyze



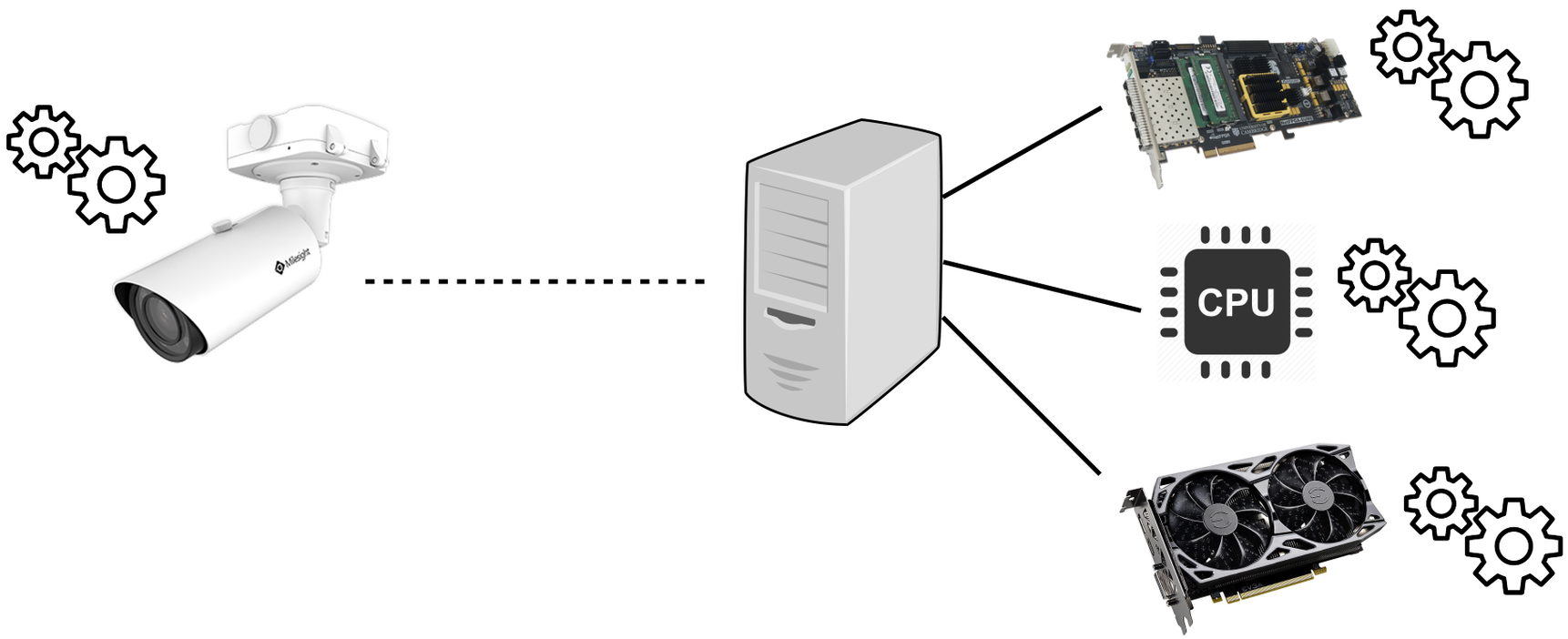
Shut down

Analyze

- **Challenging to achieve on a distributed platform!**

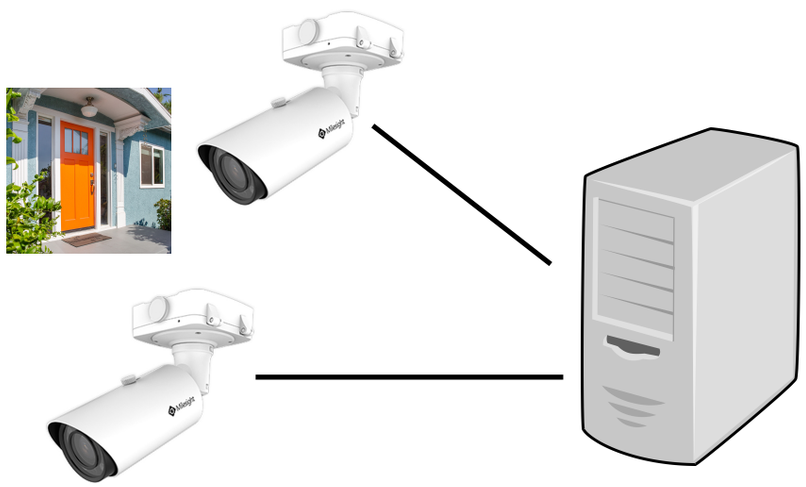
Three Fundamental Challenges for Distributed Real Time IoT Analytics

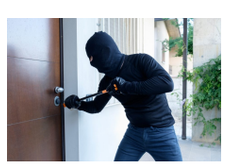
- Challenge 1: At the Application Layer
  - Compute fabric is both *distributed* and *heterogeneous*
  - Resource availability changes dynamically
  - Challenging to map analytic computations to compute fabric



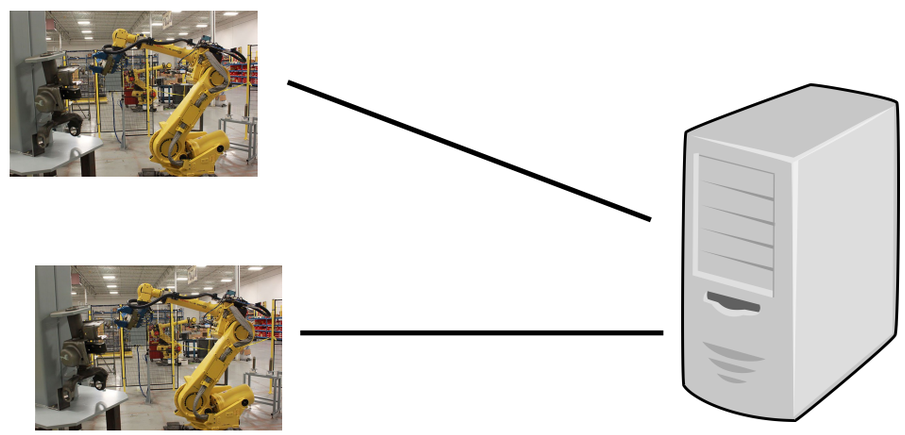
Three Fundamental Challenges for Distributed Real Time IoT Analytics

- Challenge 2: At the Network Layer
  - Shared communication channel between sensor network streams
  - Streams have different (changing) bandwidth, latency, reliability req
  - Need *intelligent* and *adaptive* network fabric





high bandwidth, low latency



high reliability

Three Fundamental Challenges for Distributed Real Time IoT Analytics

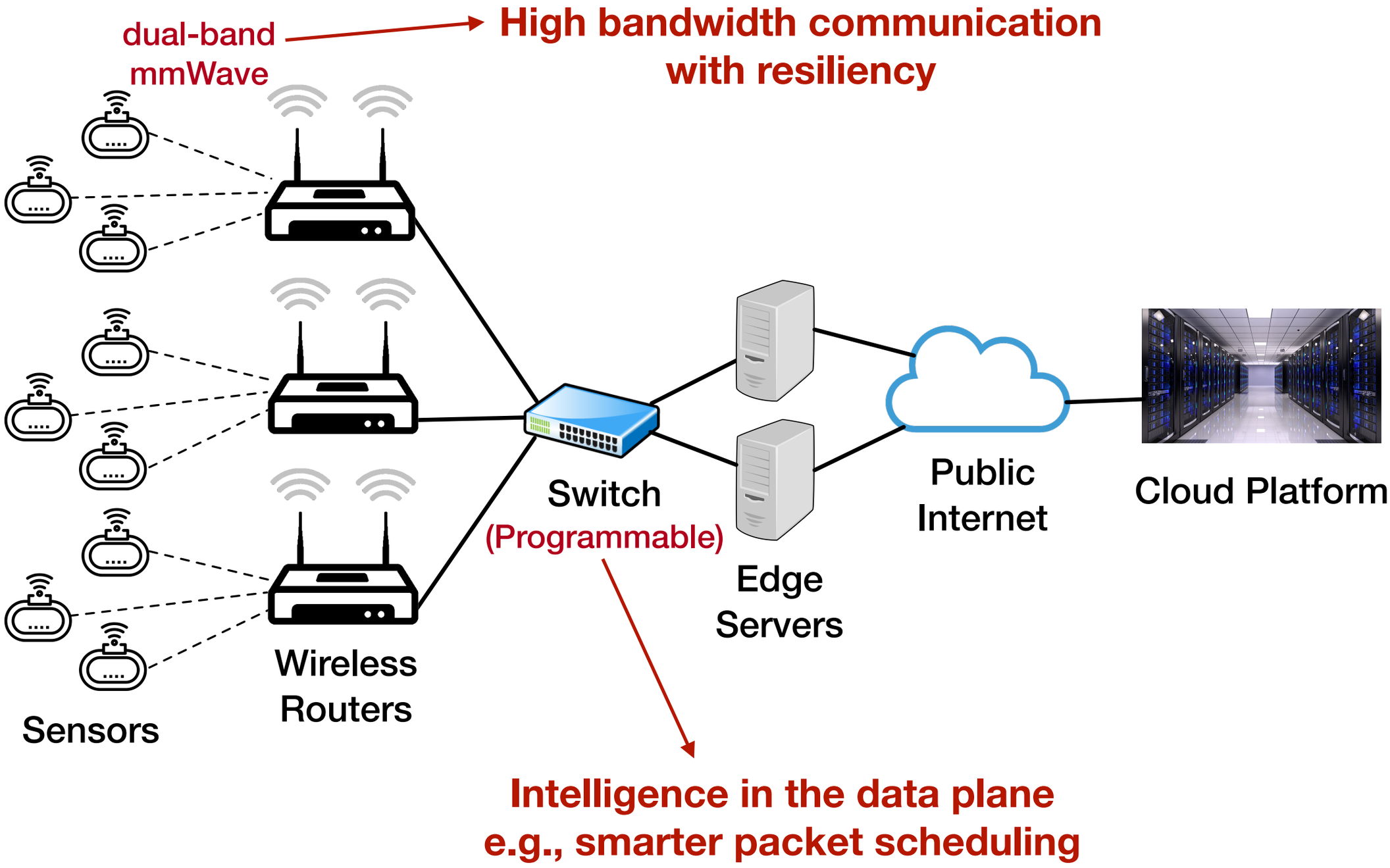
- Challenge 3: At the Physical Link Layer
  - High analytics accuracy requires multi-Gbps data communication
    - Traditional architectures use WiFi or Bluetooth
    - …very limited bandwidth!

Our Approach: RAMPS

- **Leverage emerging networking technologies**
  - Dual-band mmWave Wireless (WiGig + sub-6 GHz radio)
  - Programmable Switches

- **Cross-stack optimizations**
  - tightly couple the application, network, and link layer stacks

Our Proposed Architecture



Cross-stack Optimizations

- **Application + Link Layer**
  - Can overhear neighbor communication over wireless channel
  - Applications can share state using this *implicit* channel
  - Also use it to estimate load on network fabric and edge devices
  - …can guide optimal mapping of computations to compute nodes

Cross-stack Optimizations

- **Application + Network Layer**
  - Offload certain analytic computations to programmable switches
    - low latency, line rate computations in network data plane

Cross-stack Optimizations

- **Network + Link Layer**
  - Use real time bandwidth estimation techniques for mmWave
  - …to estimate real time available link layer bandwidth
  - Use it to guide scheduling decisions at transport layer

Cross-stack Optimizations

- **Application + Network + Link Layer**
  - Use application's knowledge of bandwidth, latency, reliability req
  - …to guide bandwidth allocation and scheduling
  - …at both network and link layers

Open Questions

- What analytic computations can and should be offloaded to programmable switches?

- How to optimally schedule data between multiple frequency bands in dual-band mmWave wireless?

- Can we leverage multi-hop mmWave networking to extend the communication range for bandwidth-demanding applications?

- What kind of scheduling policies should be implemented in programmable switches to meet bandwidth, latency, reliability requirements of all network streams?

Thank you!