Beyond the Cloud: Edge Computing

Tessema Mengistu (Ph.D.)

mengistu@cs.vt.edu

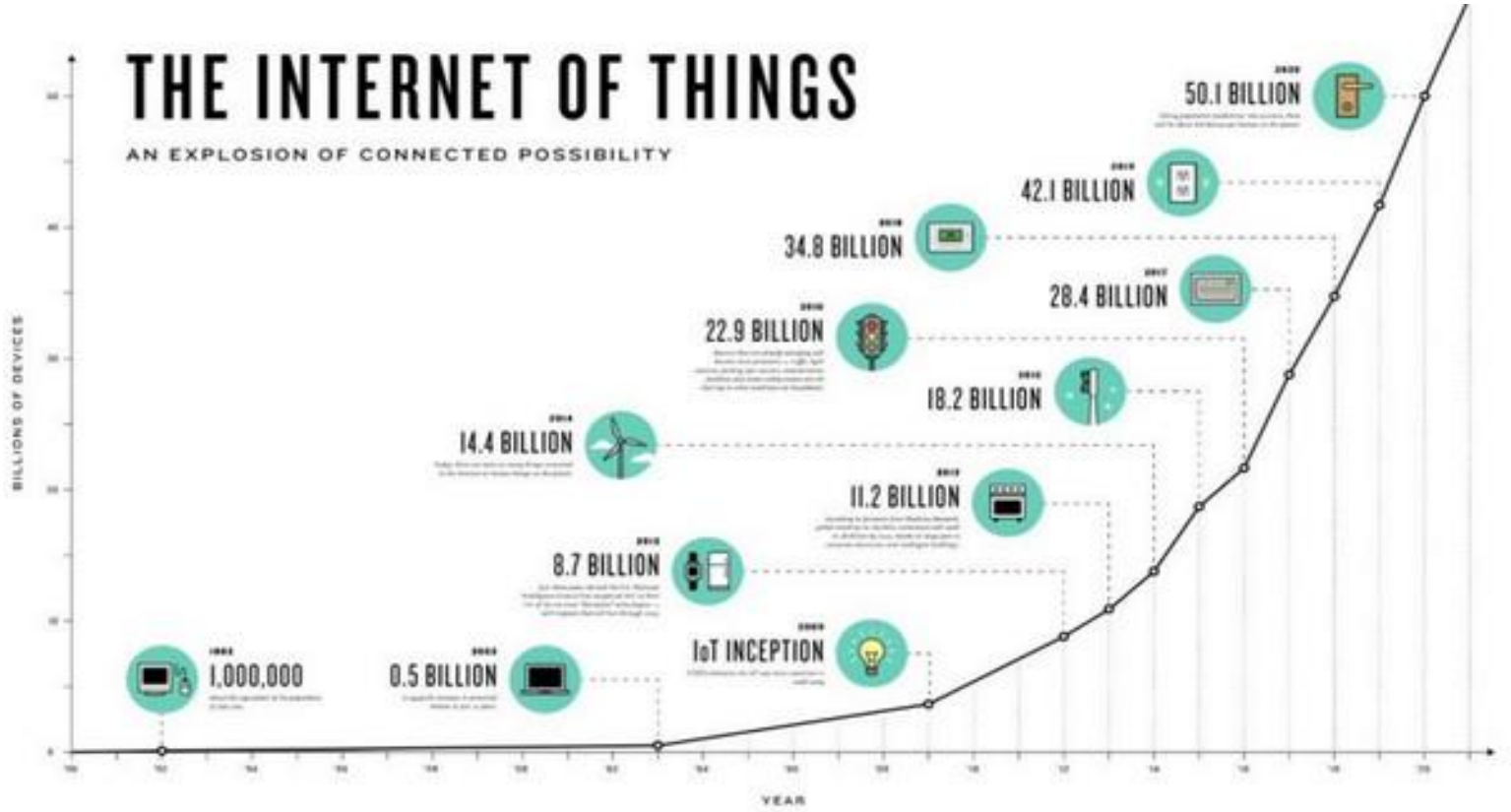
Outline

- Current Computing Landscape
- Edge Computing: An overview
- Edge Services

Current Computing Landscape

- Machine intelligence is becoming more prevalent
  - Powerful machine learning algorithms such as deep learning
  - Generative AI – Large Language Models
- Proliferation of IoT devices
  - The number of wireless-connected IoT devices passed the 50 billion mark by 2020
  - Produces huge amount of data – zettabytes
  - Poised to provide ubiquitous intelligence
  - Limitation in processing and storage powers

Current Computing Landscape



Cisco Systems Inc.

Current Computing Landscape

Conventional Cloud

- Cloud Computing has been catering the resource demands of most application
- "Data Center" approach.
  - Contains tens (if not hundreds) of thousands of servers.
- Utility computing vision of Cloud Computing is almost achieved.

Source: https://www.siebeswart.nl/image/I0000hjK4X4MnxPA

Current Computing Landscape

- Problems with processing in the Cloud
  - Latency
    - Large round trip network latency
  - Data gravity
    - Moving large amount of data is expensive
  - Data velocity
    - Realtime application
  - Bandwidth bottleneck
  - Privacy
    - Sensitive user data
  - Lack of context

Current Computing Landscape

- One of the solutions to the problem is moving the cloud infrastructure closer to users/devices by creating mini data centers or using devices at the edge of a network
  - **Edge Computing**

Edge Computing: An overview

- Content Delivery Networks (CDNs)
  - Precursor to Edge Computing
  - Offers fast and reliable content delivery and reduce communication bandwidth by caching and replication
  - A CDN receives the content from an origin server, then replicates it to its Edge cache servers
    - The content is delivered to an end-user from the “closest” Edge server

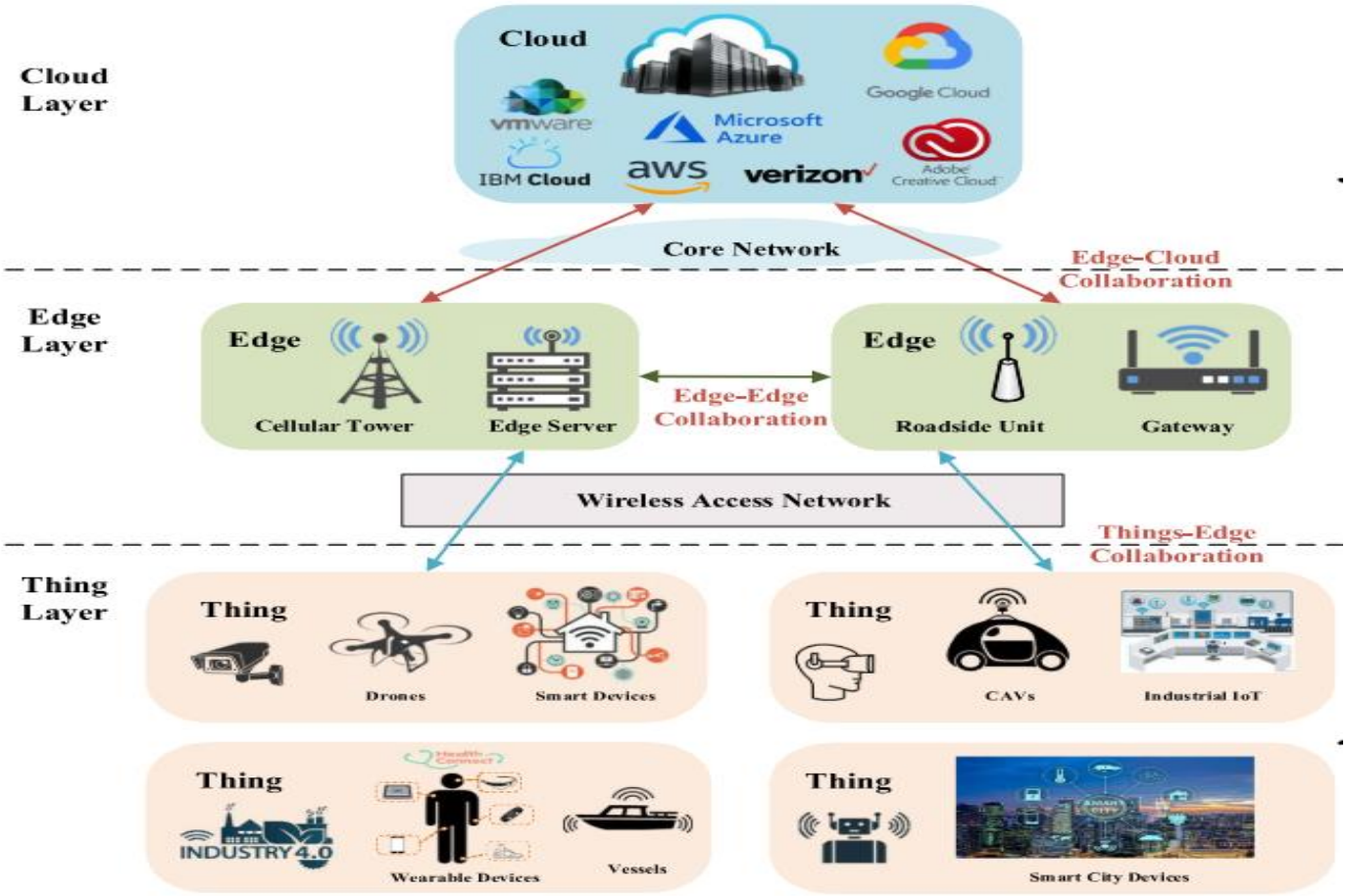
Edge Computing: An overview

- Edge Computing
  - An enabling technology allowing computation to be performed at the edge of the network, on downstream data on behalf of cloud services, and on upstream data on behalf of IoT services
  - Offload the data processing, storage, and computing operations from the cloud to the edge of the network
  - Can provide better quality-of-experience (QoE) and quality-of-service (QoS)

Edge Computing: An overview

- Components:
  - Perception/Things Layer
    - Edge Devices
    - The edge device has limited compute/storage resources
      - IoT, smartphones, etc.
  - Edge Layer
    - Edge infrastructure
      - Located directly adjacent to the access or last mile network
        - Example: cellular base stations
  - Cloud Layer
    - Cloud infrastructure

Edge Computing: An overview



Source: Resource Scheduling in Edge Computing: A Survey

Edge Computing: An overview

- There is no standardized specification about the edge of a network and the devices that are expected to participate in the edge vary
  - Vehicles , cellular base stations, networking devices , cloudlets, smart-phones, etc.

Edge Computing: An overview

- Cloudlet
  - A trusted, resource-rich computer or cluster of computers that are well-connected to the Internet and available to nearby mobile devices (Wi-Fi enabled)
  - A three-tier architecture - Mobile Device-Cloudlet-Cloud
  - The Open Edge computing initiative was also evolved from the Cloudlet project
- Multi-access Edge Computing (MEC)
  - The establishment of edge servers between the cloud server and edge devices for offloading computations
  - Embedded in mobile networks

Edge Computing: An overview

- Enabling technologies
  - Advances in wireless networks
    - 5G, Zegbee, Bluetooth, etc.
  - Datacenter technology
    - Modular datacenters

Edge Computing: An overview

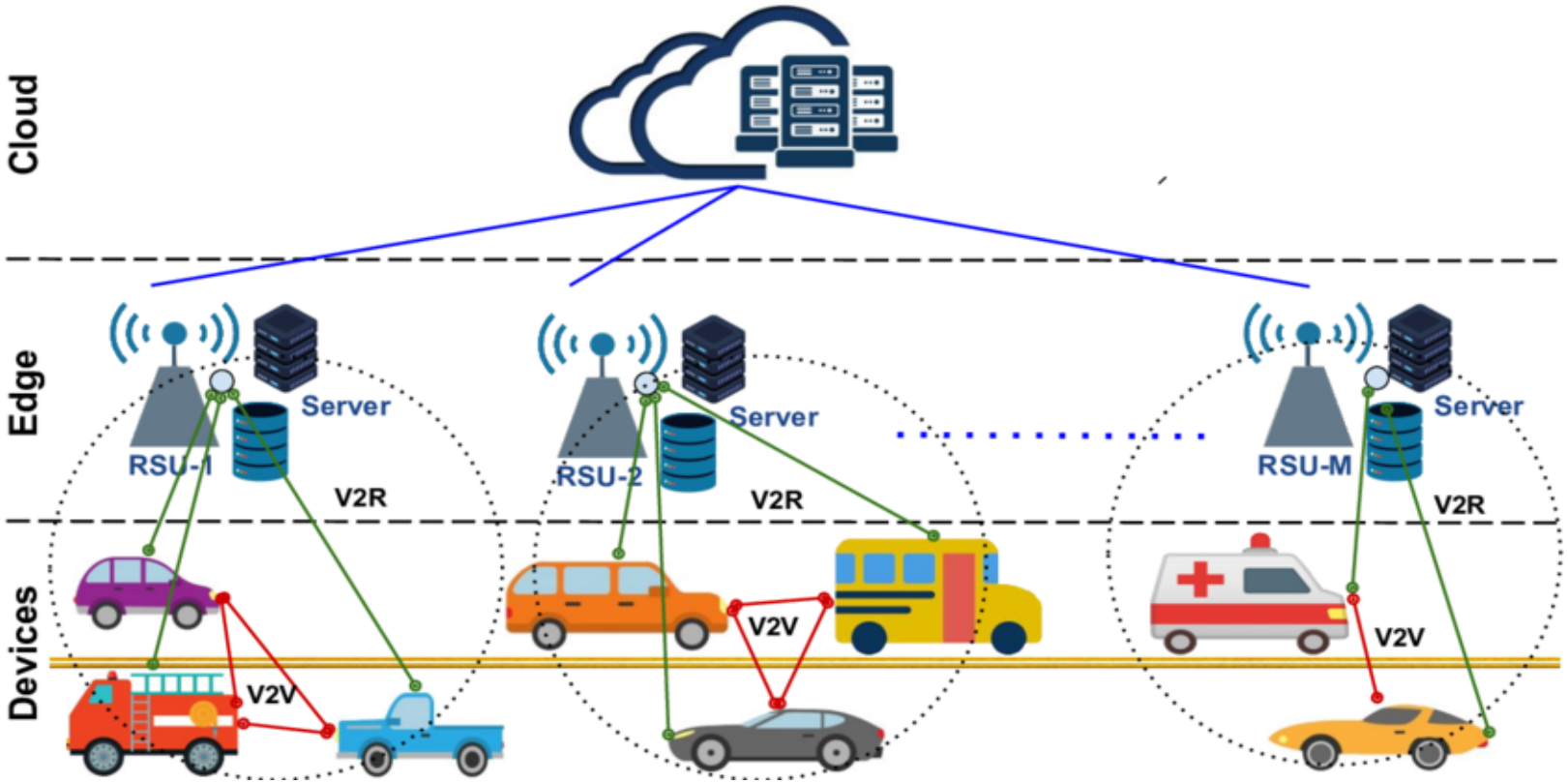
- Advantages of Edge Computing
  - Reduction in computation latency to meet the customer demands
  - Reduction in network bandwidth
  - Preservation of data security and privacy
  - Good reliability even on network failures
  - Reduction in operational costs due to communication, storage, and processing
  - Better data governance due to improvement in quality and usability of data

Edge Computing: An overview

- Limitation of Edge Computing
  - Limited computational and storage resources
    - Compare to Cloud Computing
  - Limited scalability

Edge Computing: An overview

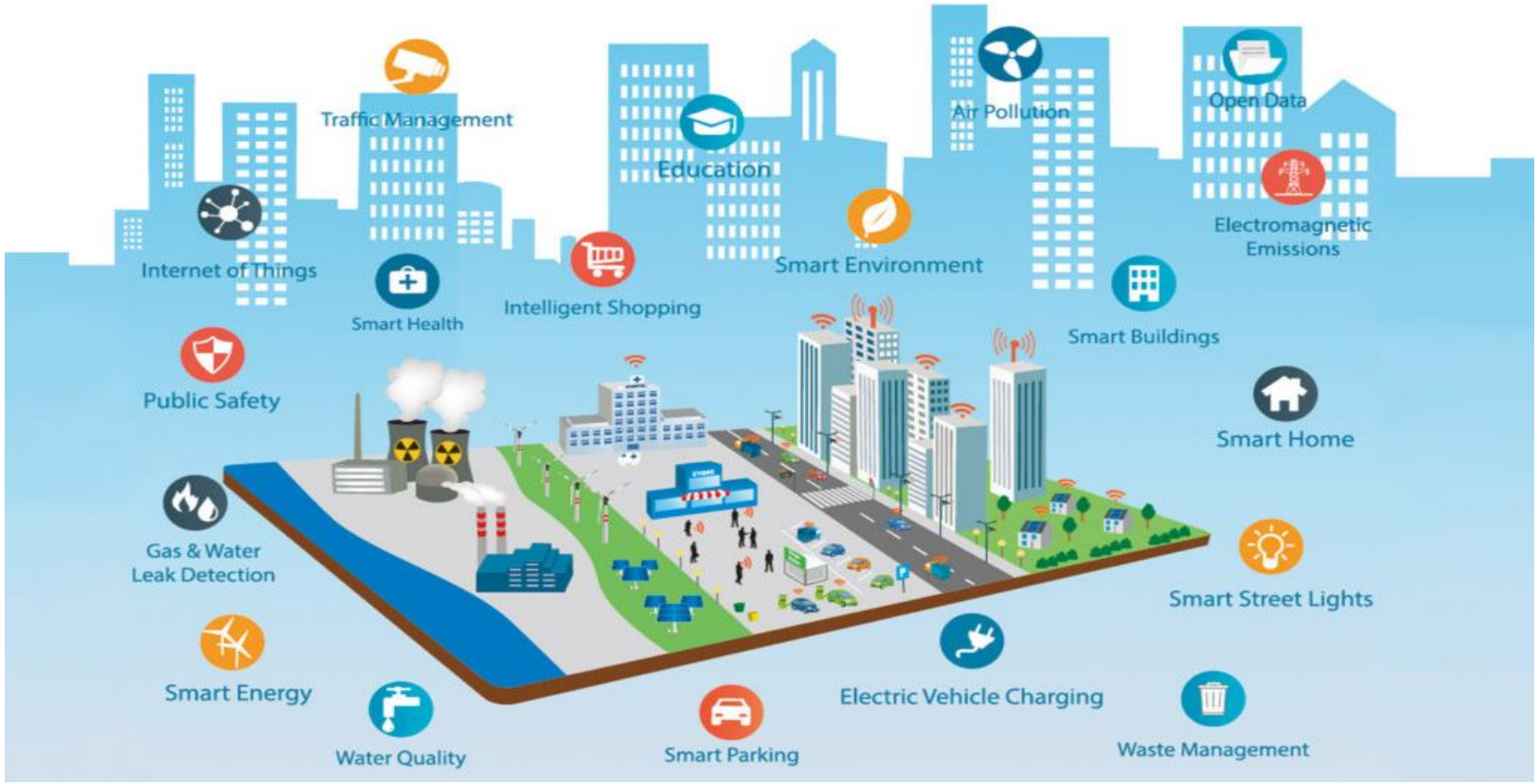
- Autonomous Vehicles



Source: DOI:10.36227/techrxiv.20115317.v1

Edge Computing: An overview

- Smart Cities



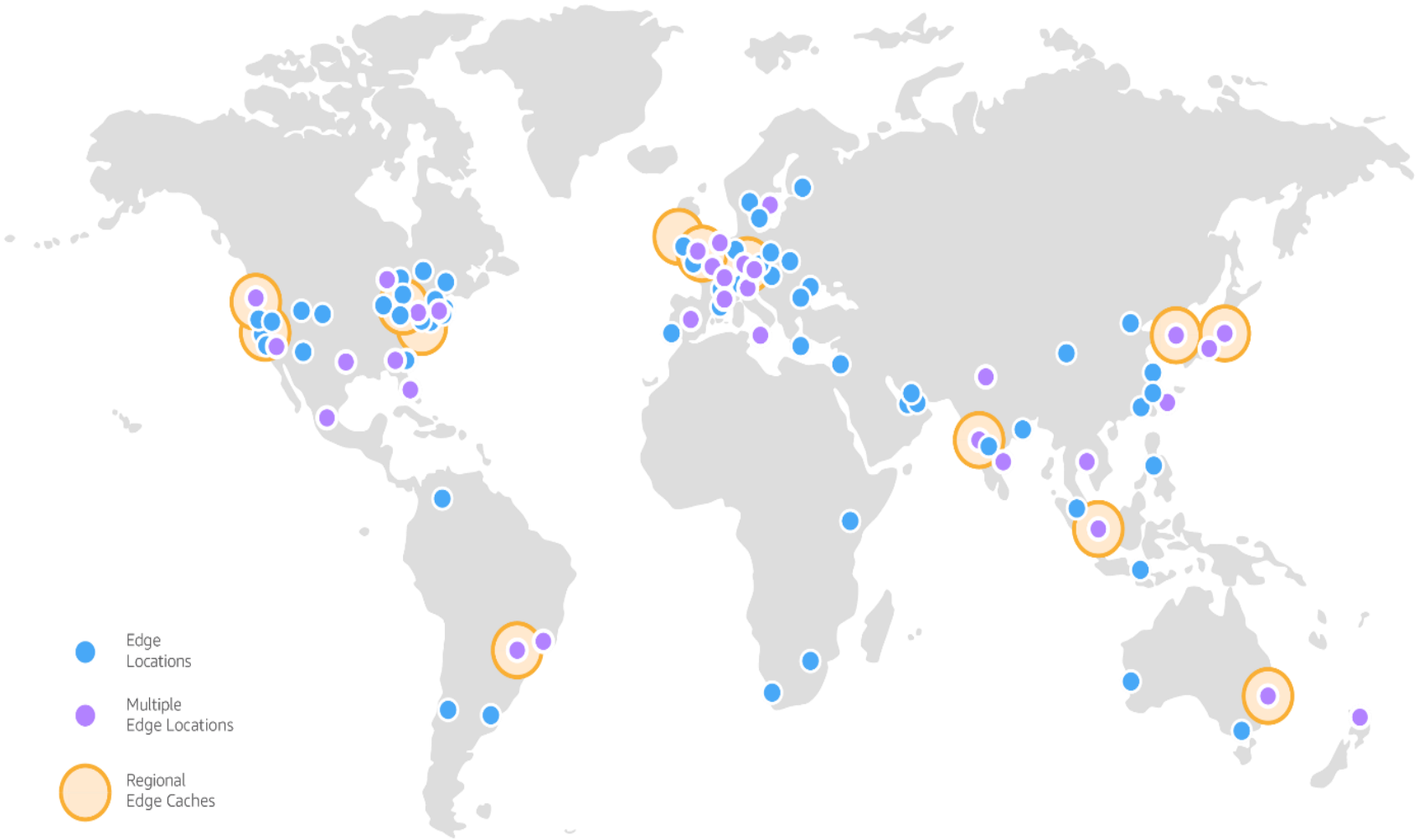
Source: Internet

Edge Services

- GCP
  - Google Distributed Cloud
    - Google Distributed Cloud Edge
    - Edge Appliance
      - Data collection, analytics, and processing
- Azure
  - Azure IoT Edge
    - Extend cloud intelligence and analytics to edge devices
  - Azure Stack Edge
    - Bring Azure compute, storage, and intelligence to the edge with Azure-managed devices

Edge Services

- Point of Presence (PoP)
  - Amazon’s Edge infrastructure
  - Hosts Amazon CloudFront, a content delivery network (CDN); Amazon Route 53, a public Domain Name System (DNS) resolution service; and AWS Global Accelerator (AGA), an edge networking optimization service



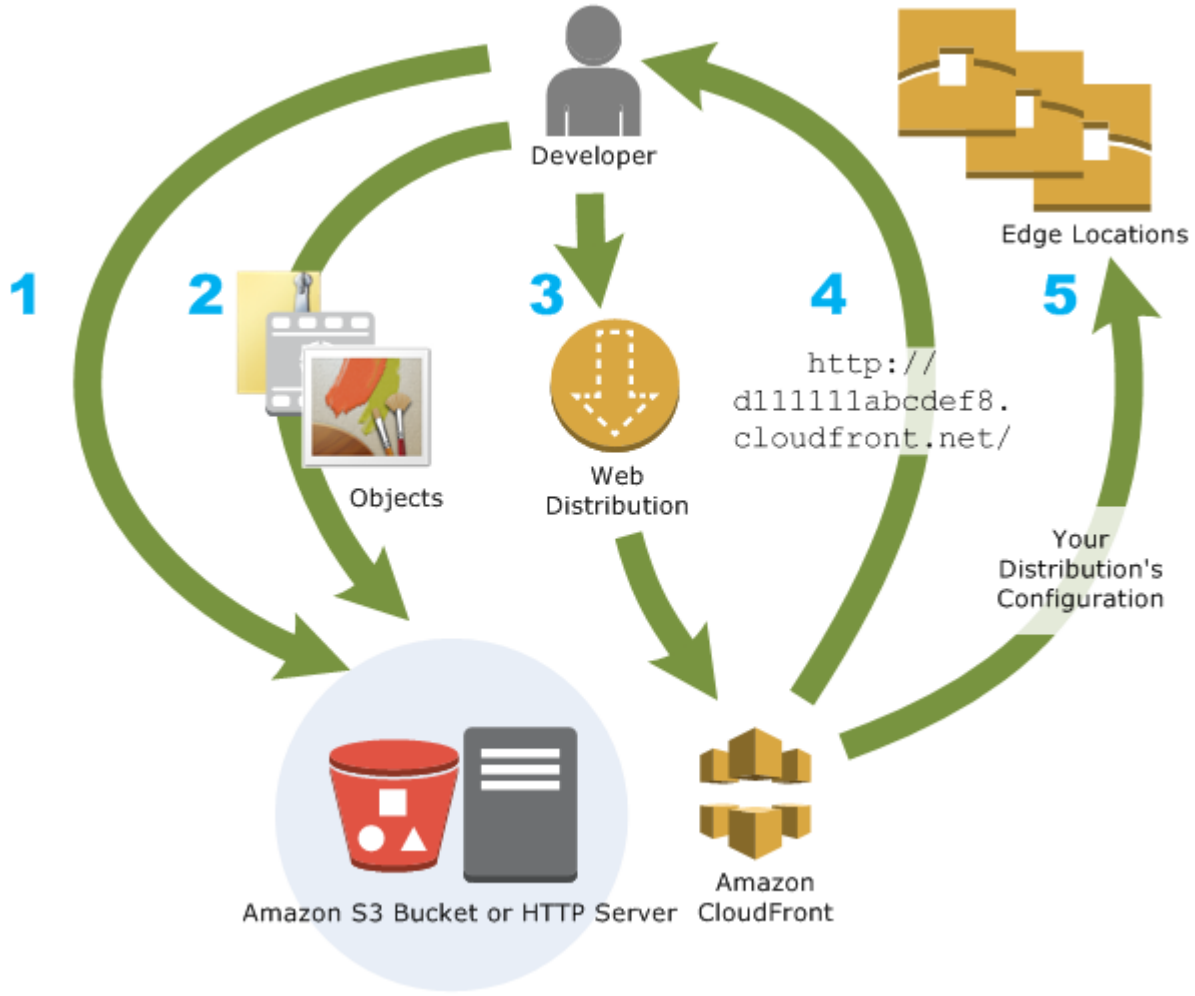
Source: https://docs.aws.amazon.com/whitepapers/latest/aws-fault-isolation-boundaries/points-of-presence.html

Edge Services

- AWS
  - AWS Outposts
    - A fully managed service that extends AWS infrastructure, AWS services, APIs, and tools to virtually any datacenter, co-location space, or on-premises facility for a truly consistent hybrid experience
  - Amazon CloudFront
    - A global content delivery network (CDN) service that securely delivers data, videos, applications, and APIs

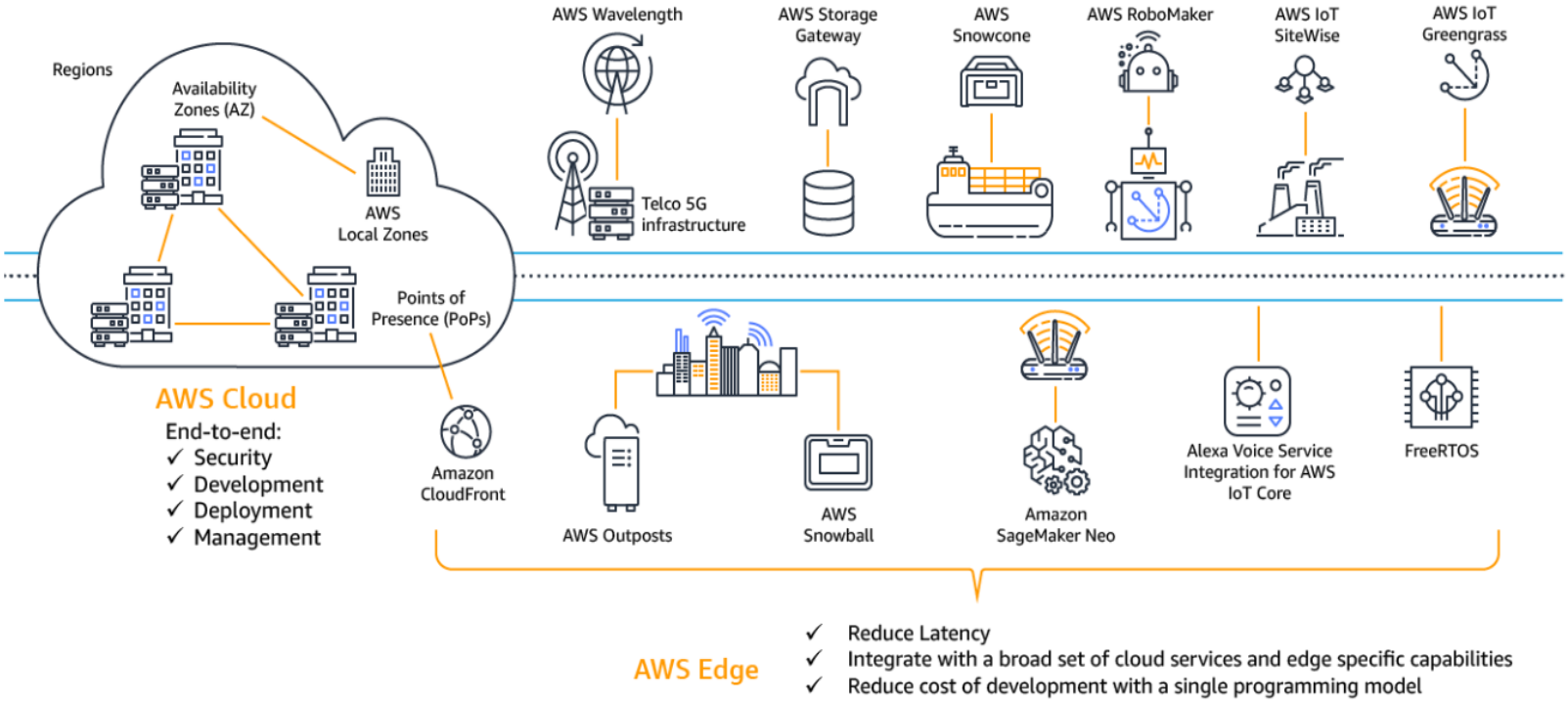
Edge Services

- AWS Edge Services
  - CloudFront
    - A web service that speeds up the distribution of static and dynamic web content, such as .html, .css, .js, and image files, to users
    - CloudFront distribution should be created to tell CloudFront:
      - Where you want content to be delivered from
      - The details about how to track and manage content delivery



Source: https://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/Introduction.html

AWS Edge Computing Portfolio



Source: https://d1.awsstatic.com/IoT/IDC-AWS-at-the-Edge-White-Paper.pdf

References

- Understanding Infrastructure Edge Computing: concepts, technologies, and considerations. Alex Marcham. 2021
- https://aws.amazon.com/what-is/edge-computing/