

Computing at the edge



draft-nordmark-t2trg-computing-edge-00

Erik Nordmark

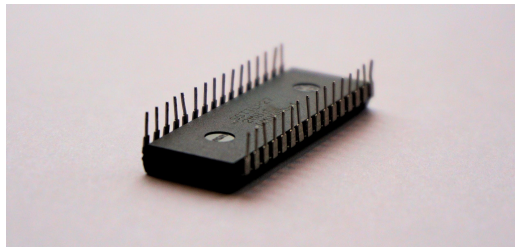
Zededa

IETF-103

Introduction

- Explore edge computing from computing perspective
 - Compute/application needs drive network needs
- Discuss engineering vs. research topics
- Consider different applications
 - Move existing applications from the cloud to the edge
 - Develop and deploy new applications
 - Applications might not be edge aware

What is computing?



- CPU/GPU, memory, I/O, optional persistent local storage
- Post deployment flexibility vs. fixed function?
- Size? Does it fit in class 2 devices?
 - Deploying existing applications vs. new applications
- Known utility cycle and power consumption?
- Is “gateway” term useful or confusing?
 - Device to gateway pattern [RFC7452] vs. product category
 - Gateway as a security/identity locus vs. translator/interop
 - General compute bricks at the edge

What is edge?



- Different notions of edge in industry
 - Cloud/Datacenter edge (not in scope)
 - Smaller datacenters “closer” (not in scope)
 - Internet edge e.g., where CDNs play
 - Provider/Telco edge
 - Enterprise/on-prem/CPE
- Where the actions are to be taken based on the analyzed data [Industry 4.0]

Why edge computing?

- Latency/real-time
- Cost and availability of bandwidth
- Autonomy/resiliency/survivability
 - Continues to function even if “uplink” connectivity gone

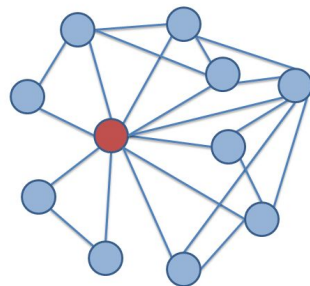


Different attributes at the edge

- Less physical security
- Different network security - no firewall to Internet
- (Physical) location is key in many cases
- Total scale much larger than the cloud datacenters
- Less network transparency - “internal” NATs and firewalls
- Multiple uplink networking technologies; redundant uplinks; mobile
- Specific downlink connectivity (6lo, BTLE, RS485, etc)
- Intermittent connectivity
- No fallback network or remote power control to handle software (updates) gone wrong

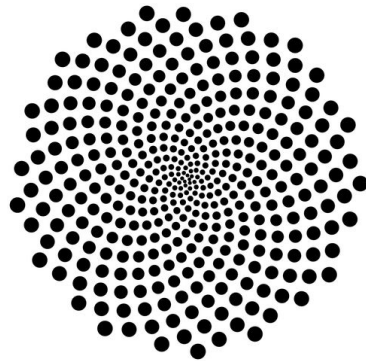


Networking for Edge Computing



- Existing cloud apps not aware of edge attributes
 - Have virtual Ethernet and part of some security group
 - Don't require them to become more aware
- Might not want to hardcode new app to use e.g., LTE
- Thus hide uplink and downlink diversity
- Provide something similar to security groups at edge

Application networking patterns



- Based on RFC7452
- Device-to-cloud - doesn't address edge motivations
- Device-to-gateway
 - Assume network topology built to match app topology?
 - Or decouple and discover gateway/compute brick?
- Topology-specific device-to-device
 - Same L2 network
- Topology-independent D2D
 - Fits (micro) service architecture at the edge

Engineering items?

- Single asset owner/manager case is mostly engineering. Range from
 - Light fixture associated with (single) light switch, to
 - GE managing all their locomotives
- Multi-tenancy is understood from cloud
 - Lease resources at the edge
- Attestation using root of trust underway [RATS BoF]
- Security and privacy patterns might be lacking
 - But perhaps not a T2TRG item



Research topics?

- Ad-hoc application patterns?
 - “I’d like a GPU near me”
 - Could be provided by some other asset owner
 - Application or the orchestration system could request
- Requires mechanisms and policies for
 - Discovery
 - Compensation
 - Security and privacy
 - Trust-worthy computing



Discus(s)?

