Rethinking Cellular Architecture and Protocols for IoT Communication

K. K. Ramakrishnan
Dept. of Computer Science and Engg.
University of California, Riverside
(collaboration with students at UCR, Prof. Koushik Kar (RPI) and Zubair Shafiq (Univ. of Iowa))
Networking Technologies for IoT

• While we depended on wired networks (even at low bandwidth) for communicating to and from sensors, wireless has by-and-large displaced them as the predominant access communication link

• In-building communication over WiFi, Zigbee, Bluetooth

• But integration of cellular communication is critical
  – Large number of applications – Health, Agriculture, Military, Utility and Environment monitoring all require external connectivity etc.
  – Focus of my talk (and of my collaboration with a small team – Prof. Koushik Kar (RPI) and Prof. Zubair Shafiq (Univ. of Iowa))

• With IoT Communications, there is a need to support a very large number of end-points – large # of communication contexts to maintain
Cellular Network Architecture (3G & LTE)

3G flow (User to Internet)
UE <-> nodeB <-> RNC <-> SGSN <-> GGSN <-> Internet

4G flow (User to Internet)
UE <-> eNodeB <-> SGW <-> PDN <-> Internet
How complex is the network environment?

Elegant layered model

Application
Transport
Network
Data Link
Physical

Reality

“Web”
HTTP
TCP
IP
UDP
GTP
MPLS
MPLS VPN
SONET/SDH/OTN
Ethernet VLAN
Ethernet PHY
DWDM
Fiber

Video
RTP

Reality
LTE Infrastructure: use of GTP Tunnels

Communication between the mobile device and IP network over tunnels

- UE: user equipment
- eNodeB: base station
- S-GW: serving gateway
- P-GW: packet data network gateway
- MME: mobility management entity
- HSS: home subscriber server
- PCRF: policy charging and rule function

Cellular Core Network

MME/PCRF/HSS

Internet and Other IP Networks

UE 1

UE 2

eNodeB 1

S-GW 1

P-GW

S-GW 2

GTP Tunnels

eNodeB 2

eNodeB 3

UE 2

Internet and Other IP Networks

UE 1
Difficulties of Tunnels: Amount of State

• Each end-point has a distinct tunnel established from eNodeB (base station) to the S-P Gateway
  • Can be a significant impediment to scalability
• Minimize state relative to # of endpoints
  • Cost is significant concern for the scale expected with IoT Communication
  • Cost of state (memory, compute) relative to revenue – especially when supporting individual nodes for infrequent communication.
Mobility Management

• Handoff without change of S-P GW – (S1 handoff)
• Results in up to **33 control messages** in total across S-GW, MME and eNBs.
• Handoff with change of S-GW or MME has more overhead
• Mobility for large # of IoT devices – overhead is of concern
More Difficulties of Tunnels for IoT support

• Having a tunnel per IoT device can also have significant impact on performance
  • Control plane messages interact at the S-GW with data plane
  • Performance penalty on control and data planes (at S-GW) can be significant as the number of tunnels grows
  • Each tunnel event (mobility/handover, attachment and to a lesser extent RRC state changes) interacts with data plane packet processing – resulting penalty can severely degrade packet throughput

• Implications particularly for IoT support: tunnel establishment (which can happen each time a sensor ‘wakes up’) – significantly degrades performance

• Low volume, infrequent traffic from IoT devices also causes the RRC state change to occur frequently – generating control messages on the tunnel
Device Power Consumption: Radio Resource Management (LTE)

- End-device runs Radio Resource Control (RRC) state machine
- Two major states: IDLE, CONNECTED
- Discontinuous reception (DRX): monitor one sub-frame per DRX cycle; receiver sleeps in other sub-frames
Establishment of Tunnel on State Transitions

- Paging
- If S-GW receives a packet to a UE in IDLE state, inform MME
- MME pages UE through base station
- Results in **15 to 19 control messages** between S-P GW, MME and eNB
Redesign for 5G with Software-based EPC

• Move to Virtualized EPC (e.g., software based S-P GW) to address scalability – common now
• But, this offers us an opportunity to re-think the protocols we use in cellular networks
  • Not just implement the same set of protocols in software
• Software EPC designed to support large numbers of eNBs (e.g., a region, metro area)
• Scale can be our friend
• Potential to completely do away with all the mobility-related complex tunnel establishment/tear-down
  • Optimize the common case
• Exploit ICN capabilities for coarser mobility