Rethinking Cellular Architecture and Protocols for IoT Communication

K. K. Ramakrishnan

Dept. of Computer Science and Engg., University of California, Riverside

Abstract: Communication for 'Internet-of-Things' (IoT) is becoming a significant component to be supported by the network infrastructures. With 'smart' sensing devices becoming ubiquitous, there is a need to support IoT communication at large scale over cellular networks. 5G is seeking to change the cellular architecture at many levels, but this also brings multiple challenges. The heterogeneity and the mix of macro and small cells exacerbate the problems introduced by mobility and hand-offs. My talk will discuss some of the challenges for future cellular networks to provide support for an environment that is critically dependent on smart sensing devices.

Introduction

Our lives have changed significantly over the last few decades with the availability of ubiquitous network connectivity, especially as wireless connected devices have been integrated seamlessly into our 'connected life'. With sensors also becoming ubiquitous, a wide spectrum of networked applications, such as Tele-health, shipping and logistics, utility and environmental monitoring, point-of-sale terminals, industrial automation, and asset tracking are growing. The need to support communication by such devices at large-scale has become a reality. Cellular networks have become the way much of the world communicates, in this current predominantly wireless world. IoT traffic will also need to be supported over such cellular networks. It is estimated that there are already tens of millions of such devices connected over cellular networks worldwide, and within the next few years this number will grow to billions [1], [2] Thus, sensor devices and smartphones will end up share the same network infrastructure. But current cellular data networks are primarily designed, engineered, and managed for smartphone usage, and the typical interactive applications used by people. The protocol framework to support IP traffic has evolved from the cellular network's initial circuit-switched/admission-controlled protocol framework developed to support voice traffic. Tunnels are established between the user device (often termed UE for user equipment) and the cellular network's network data center for each session, which is a significant overhead unsustainable for typical device to device communication. Given that the population of cellular-based IoT devices may soon eclipse that of smartphones, scalable support for these devices in cellular networks is arguably the biggest challenge that we face towards making vast numbers of IoT devices accessible over the Internet.

Previous studies, such as [3], of IoT traffic over cellular networks has shown that such traffic has distinct characteristics that set it apart from traditional smartphone traffic. For instance, IoT devices generate much less traffic on a per-device basis as compared to smartphones. They have a much larger ratio of uplink to downlink traffic volume compared to smartphones. We may have to consider new cellular architectures and protocols to avoid contention between low volume, uplink-heavy IoT traffic and high volume, downlink-heavy smartphone traffic. IoT devices, with a few exceptions, are less mobile than smartphones. Thus, careful network resource allocation is required to avoid contention between low-volume IoT traffic and high-volume smartphone traffic. All of these suggest a need to re-think the architecture of cellular networks to support IoT communications.

Limitations of Existing Solutions

The 3G-PP architecture that is currently used for mobility handling in cellular networks - which is centered around the concept of "tunnels" - is inefficient, and clearly would not scale, even if a small fraction of all IoT devices are mobile and have to be addressable and reachable through the Internet. As we move to a predominantly IP-based data network, the protocol structures and layering that is used to support IP-based data traffic have become excessively complex. A significant component of that complexity comes from the use of tunnels for the communication between the mobile device and the Network Data Center (NDC). In fact, even for sending IP packets over to the Internet, and potentially to a server that is close (in terms of IP hops) to the mobile device, all of the packets are carried in a tunnel that traverses the cellular network's 'packet core network' (that includes the backhaul network from the cellular base station all the way to the NDC). The NDC is involved in a variety of functions including assigning local IP addresses (DHCP), providing the network address translation (NAT) functionality and terminating the tunnel. The number of protocol headers that encapsulate the packet can be excessive, adding overhead; further, there are multiple routing and load balancing algorithms that get involved. All of this adds complexity and state in the network. Since this complexity and state maintenance is on a per-mobile device, the processing, and the cost of maintaining this state becomes a significant burden.

As we evolve the cellular network to the next generation, it is desirable to evolve from this approach of being centered on the concept of 'tunnels'. Furthermore the traditional approach of using Mobile IP to deal with the mobility of devices also involves substantial overhead. Mobile IP requires an anchor point (again, this is typically at the NDC in a cellular network). It results in poor routing of traffic, adding unnecessary latency. It is desirable to evolve away from this approach. Some of the efforts in the IETF, such as LISP-Mobile Node [4] and work in the Distributed Mobility Management working group (e.g., [5]), and current ongoing projects funded by the NSF in the Future Internet Architecture program [6], are directed at improving this. The vision that IoT communication is the next big challenge for cellular wireless communications, as cellular traffic is dominated by machine-to-machine (M2M) or IoT communication (especially short messages requiring low latency).

Potential Solution Directions

A careful examination and optimization of tradeoffs between the different end-system and in-network functions associated with IoT communication naturally leads us to re-examine and re-design the core operations and protocols at the *link*, *network* and *transport* layers – at each layer individually, but also in harmony with the other layers so that the overall system operates efficiently.

At the link layer, it would be highly desirable to revisit cellular RRC protocol state machine design and parameter optimization for the support of IoT communications. Given the widely varying characteristics of IoT traffic, it is desirable for the transport layer to provide a dynamic and adaptive set of core components to support such traffic. Some of the choices to be examined are to selectively introduce reliability and in-sequence delivery, flow and congestion control as well as seamless adaptation to mobility as needed. *Information Centric Networking (ICN)* is another promising direction that can be beneficial to support the needs of mobile IoT devices. Several projects supported by the US National Science Foundation for the Future Internet Architecture (FIA) seek to improve the support for the current needs of information dissemination and end-system characteristics [7, 8]. One of the aspects of the MobilityFirst FIA[8] is to provide better support of mobility through the use of a name rather than location for communication. Furthermore, capabilities such as publish/subscribe for information generated by IoT devices are more naturally and efficiently supported in ICN [9]. The network layer enables access to information and devices by name. By using late-binding to do the name resolution, we can continue to correct the routing as the PDU progresses towards the ultimate destination, thus supporting mobility seamlessly [8].

We are currently exploring these directions for enhancing networking support for IoT, with particular emphasis on how cellular communications can seamlessly support very large-scale IoT environments.

References

1. 3G machine-to-machine (M2M) communications: Cellular 3G, WiMAX, and municipal Wi-Fi for M2M applications. Technical report, ABI Research, 2007.
2. The global wireless M2M market. Technical report, Berg Insight, December 2010.
3. M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, and J. Wang. A First Look at Cellular Machine-to-Machine Traffic - Large Scale Measurement and Characterization. In ACM SIGMETRICS, 2012.
4. D. Farinacci, D. Lewis, D. Meyer, and C. White. LISP Mobile Node. Internet-Draft, Internet Engineering Task Force (IETF), July 2014.
5. D. Liu, J. Zuniga, P. Seite, H. Chan, and C. Bernardos. Distributed Mobility Management: Current practices and gap analysis. Internet-Draft , Internet Engineering Task Force (IETF), September 2014.
6. NSF Future Internet Architecture Project. http: //www.nets-fia.net.
7. L. Zhang and et al. Named data networking. SIGCOMM Comput. Commun. Rev., 44(3), July 2014.
8. D. Raychaudhuri, K. Nagaraja, and A. Venkataramani. Mobilityfirst: A Robust and Trustworthy Mobility-centric Architecture for the Future Internet. SIGMOBILE Mob. Comput. Commun. Rev., 16(3), December 2012.

9. J. Chen, M. Arumaithurai, L. Jiao, X. Fu, and K. K. Ramakrishnan, COPSS: An efficient content oriented publish/subscribe system. In ACM/IEEE ANCS, October 2011.