"Infrastructure for the Personal Cloud"

Joseph Cavallaro, ECE Dept. Rice Univ. cavallar@rice.edu and Christoph Studer, School of ECE, Cornell Univ. studer@cornell.edu

Vision: As we move beyond the internet to a global interconnection of users and devices in the billions, new challenges emerge in communication, networking, and computation. We envision a scalable communication infrastructure that will require orders of magnitude increases in power and bandwidth. This infrastructure will move away from a rigid centralized system into a more decentralized collection of resources. This decentralized system will actually be tightly connected to not only data transfer but also to local computing and storage — that will lead to a "personal cloud" computing network. This "personal cloud" will be more agile and able to adapt to dynamic environments and mobility. We can see the beginnings of this trend in recent concepts and literature in the topics of IoT, "fog" computing, smart cities, tactile internet, etc. All of these trends build on the needs for fast local communication and computation for quick response, for example accident avoidance in autonomous vehicles, yet with rich access to more traditional big data and "cloud services" such as navigation guidance and location aware services.

In this environment, we need to identify what is nearby (via localization) in terms of network access, and then pick the most appropriate communication strategy, perhaps in a software defined radio (SDR) capability. In order to respond quickly, for example urgent real-time control in an industrial setting, and for data logging sensor data in a maintenance database, a system (or perhaps a human) will need to quickly decide whether it is more beneficial to go directly to a large traditional base station or whether a short hop to a smart local gateway is sufficient.

Grand Challenges: There are many challenges in developing a "personal cloud" to provide this level of fast, yet energy and spectrum efficient wireless network and computing services. For instance, new scalable standards with more agility in SDR will be required. Also, what type of computation substrate (FPGA or processors) will provide the greatest network and computing services? The goal is to provide the best efficiency so that local edge computing can be provided when real-time analysis is needed and fast communication with a cloud server can be provided when data science applications are active.

This grand challenge can be in the form of designing a single standard that covers these diverse scenarios, without having a plethora of different protocols and modulation schemes, i.e., a single standard that scales from ultra low power transmission to high throughput Gb/s transmission. It would be inefficient for a person or an IoT cluster to continuously switch between standards. We would envision the development of a unified standard that covers orders of magnitude in terms of bandwidth, energy efficiency, latency, range, etc.

Infrastructure and Testbeds: Design space exploration will be needed to assess the area-time-power computation and communication efficiencies of new algorithms and architectures for a "personal cloud." Experience can be gained from the FPGA and programmable architectures and testbeds developed in universities and industry for 3G and 4G communication systems in small scale MIMO, such as the Rice WARP wireless testbed, the NI/USRP testbeds, etc. These existing testbeds are being effectively used in 5G proposals, such as in investigation of massive MIMO (Argos at Rice, NI/USRP at Lund) but are reaching their limits due to the complexity of system integration and signal routing. A new generation of testbeds having greater modularity, flexibility, with a variety of communication, computing and storage performance at each node would be valuable in studying and identifying opportunities, bottlenecks, and design tradeoffs in a more heterogeneous and adaptable "personal cloud" network.