Research Statement
Ayon Chakraborty

The goal of my research is to improve the state of Internet connectivity for the future world. Although majority of the devices today are wireless enabled, with the exponentiating growth in the number of ‘connected’ devices\textsuperscript{1} it is unclear how incremental changes to the current network infrastructure can sustain such growth. The problem stems from the fact that the capacity of a wireless network is fundamentally capped by limited availability of physical resources, e.g., wireless spectrum. With innovations in applications like 4K video streaming or virtual reality, demands for good user experience (QoE) and an exponentiating user base: such capacity is already close to hit the wall. For instance, even though the capacity of LTE networks today is close to the Shannon limit, it is still unclear how to meet capacity demands in near future with the surge of data traffic.

Thesis Work
To have a holistic view of the problem, my dissertation research stands astride two extreme ends of the network protocol stack: (a) the Application layer and (b) the Physical layer. Essentially, the general direction of my research looks at optimizing resource utilization to improve capacity / performance in these two layers. I made contributions to understanding and characterizing Quality of Experience (QoE) of different types Internet applications (video streaming, web browsing, VoIP etc.) in heterogeneous network conditions [ CellNet ’13, IMC ’14, CoNEXT ’16 ]. Second, I contributed to augmenting wireless capacity by harvesting unused radio frequency spectrum or white spaces [ MOBICOM ’13, CoNEXT ’14, DCOSS ’16, Hot Wireless ’16, INFOCOM ’17 ]. Third, I have also worked on network-side positioning of cellular band devices [ INFOCOM ’15 ].

Many of such problems can be solved using analytical (whitebox) techniques that depend on meticulous mathematical modeling of the concerned system. However the complexity and scale of today’s systems deem such whitebox solutions to be rather impractical due to lack of availability of granular information. Inspired by the recent success of data-driven (blackbox) analysis in many fields of computing, my dissertation research explores such approaches in modeling and optimizing performance of networked systems. In collaboration with the industry, much of my research efforts translated to building actual system prototypes that were passed on to respective product teams. For instance, my work on the QoE management framework, \textit{ExBox} [ CoNEXT ’16 ], is being integrated with Hewlett-Packard/Aruba’s network selection technology. My work on network-side localization [ INFOCOM ’15 ] has been transferred to Huawei Technologies.

Characterizing Quality of Experience: Today’s mobile application ecosystem is a complicated mess. Various entangled dimensions add to such complexity: carriers, networks, client devices and of course diverse types of application programs. Amidst such complexity what matters the most is Quality of Experience (QoE) perceived by end users. Guaranteeing good QoE for the user is a challenge in mobile apps due to their diverse resource requirements and the resource-constrained, variable nature of wireless networks and mobile devices.

A key question we ask is: how to provision a network given a set of diverse applications and client devices, or how to optimize the overall user experience for multiple users and applications in a given network? We develop a user feedback-based system called \textit{Adapp} that learns the best network to use for an app when multiple networks are available [ CellNet ’13 ]. \textit{Adapp} optimizes QoE at an individual user level on the client-side. To complement this effort, we also develop a network-side system called \textit{ExBox} (\textit{Ex}perience \textit{m}iddle\textit{Box}) that uses a new experiential capacity-based formulation to aid admission control and network selection [ CoNEXT ’16 ]. \textit{ExBox} can handle multiple different types of applications in the network and dynamic network conditions. Both \textit{Adapp} and \textit{ExBox} have been tested on real network testbeds. My dissertation also contributes towards

---

\textsuperscript{1}According to Qualcomm, the number of interconnected devices will rise from 7 Billion in 2017 to about 25 Billion+ by the end of this decade resulting in about a 1000× increase in mobile data traffic.
extensive QoE measurement studies for multiple applications across different commercial network providers [IMC ’14] and hardware vendors in specific networking contexts.

**Optimizing Shared Spectrum Access Technologies:** With the explosion of mobile data, there is a growing realization that RF spectrum must be treated as an important resource that is in limited supply. Policy makers and researchers alike are promoting various forms of spectrum sharing models to improve spectrum utilization. This has promoted new spectrum sharing regimes where a diverse set of wireless technologies, including broadcast television, radar and various forms of wireless communications, co-exist in the same spectrum band while respecting specific spectrum rights. Just like any other resource with mismatched demand and supply, all steps towards better utilization have also increased the need for large scale spectrum monitoring.

Our key question is: how effectively can we monitor available spectrum opportunities across time and space? One key technology in this space is spectrum databases that store spectrum availability information. We augment existing spectrum database technologies to improve their accuracies in a cost-effective fashion [CoNEXT ’14]. The idea is to supplement existing model-based techniques with actual spectrum measurements. We also contribute towards making the spectrum measurements themselves scalable by developing techniques to perform spectrum sensing on mobile devices [MOBICOM ’13, Hot Wireless ’16]. These efforts culminate building a spectrum database system called SpecSense that can schedule and collect measurements from a distributed system of spectrum sensors in order to estimate spatio-temporal patterns in spectrum availability [DCOSS ’16, INFOCOM ’17].

**Network-Side Localization of Cellular Devices:** We address the problem of network-side localization where cellular operators are interested in localizing cellular devices by means of signal strength measurements alone [INFOCOM ’15]. While fingerprinting-based approaches have been used recently to address this problem, they require significant amount of geotagged (‘labeled’) measurement data that is expensive for the operator to collect. Our goal is to use semi-supervised and unsupervised machine learning techniques to reduce or eliminate this effort without compromising the accuracy of localization. Our experimental results in a university campus (6 sq. km) demonstrate that sub-100 m median localization accuracy is achievable with very little or no labeled data so long as enough training is possible with ‘unlabeled’ measurements. This provides an opportunity for the operator to improve the model with time.

**Broader Impact** Given the reach of cellular Internet, this research will positively impact a broad spectrum of the population. SpecSense-like infrastructure will enable several compelling applications. For example, improving mobile application’s QoE is crucial for the millions of users for whom the smartphone is their only computing device. Research on leveraging unused spectrum can potentially improve cellular bandwidth and improve connectivity; this will reduce the cost of expensive cellular data plans making mobile Internet more affordable.

**Future Directions**
My long term vision is to design scalable network infrastructure that can accommodate the exponentiating growth of connected devices and corresponding data traffic [ATC ’16]. The challenge here is managing network resources so as to provide optimal QoE. This improves overall user engagement while providing a cost-effective network. Carrying forward from my dissertation research, I am planning to focus on the following directions in the immediate term.

**Large Scale Monitoring Infrastructure:** Future applications (e.g., smart cities) will depend on a huge amount of geo-spatial measurement data of different modalities collected from a diverse array of sensors/IoT devices. Some minimal fidelity of such data needs to be guaranteed given a cost budget. For example, a system like SpecSense mentioned above needs to create an accurate spectrum map given a budget for the sensor costs. Spectrum sensors, depending on the cost, can have heterogeneous characteristics resulting in diverse cost-performance tradeoffs. This requires accurate modeling of a sensor’s performance given its characteristics (hardware/software) that impact cost.
Research Statement

Ayon Chakraborty

Second, there are no clear understanding about how to deploy such sensors in a geo-spatial sense or how to schedule them across space/time for performing sensing tasks. The latter is important as it may be impractical for a sensor to continuously make measurements due to energy limitations at individual sensors and/or excessive network traffic to collect data from them. Third, such sensor data may contribute to real time decision making in certain applications. Hence optimizing end-to-end latency is a vital requirement in such a monitoring infrastructure. My vision is to create the necessary toolbox combining algorithms, sensor systems, network protocols and big data techniques that will be an essential component of future mobile networks – 5G and beyond – supporting upcoming applications such as IoT or smart cities. To deliver on this vision, I will go beyond narrowly focused academic setting and work with industry to understand realistic applications and constraints.

Data Driven Modeling of Network Performance: As networks grow more complex, modeling, understanding and debugging network performance issues become harder. For instance, there is now little understanding whether a poor video streaming experience on a smartphone should be attributed to capacity issues in the last hop link or in the mobile network core, Internet congestion, CDN provider, the device hardware or a combination of multiple such factors. Disentangling such dependencies is not a straightforward problem in a realistic setting as majority of the network components lack sufficient visibility to the end user application so that traditional modeling approaches can be applied. I will exploit crowdsourcing techniques and data driven modeling methodologies to address such modeling problems. The idea is to collect scenario-specific data for application performance via crowd-sourcing and then use such large-scale data to draw inferences about performance dependencies. Such approaches will be critical in building mobile applications and provisioning networks in the future.

References


